JULIANA MACHADO RUGOLO

Perfil molecular das vias de sinalização dos carcinomas de pulmão de células não pequenas por Sequenciamento de Nova Geração (NGS) com ênfase à Transição Epitélio Mesenquimal (EMT)

Tese apresentada à Faculdade de Medicina da Universidade de São Paulo para obtenção do título de Doutor em Ciências

Programa de Patologia

Orientadora: Profa. Dra. Vera Luiza Capelozzi

São Paulo

2022

JULIANA MACHADO RUGOLO

Perfil molecular das vias de sinalização dos carcinomas de pulmão de células não pequenas por Sequenciamento de Nova Geração (NGS) com ênfase à Transição Epitélio Mesenquimal (EMT)

Tese apresentada à Faculdade de Medicina da Universidade de São Paulo para obtenção do título de Doutor em Ciências

Programa de Patologia
Orientadora: Profa. Dra. Vera Luiza Capelozzi

São Paulo

2022

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

Rugolo, Juliana Machado
Perfil molecular das vias de sinalização dos carcinomas de pulmão de células não pequenas por Sequenciamento de Nova Geração (NGS) com ênfase à Transição Epitélio Mesenquimal (EMT) / Juliana Machado Rugolo. -- São Paulo, 2022.
Tese(doutorado)--Faculdade de Medicina da Universidade de São Paulo.
Programa de Patologia.
Orientadora: Vera Luiza Capelozzi.

Descritores: 1.Carcinoma pulmonar de células não pequenas 2.Sequenciamento de nucleotídeos em larga escala 3.Biologia computacional 4.Medicina de precisão 5.Biomarcadores 6.Mortalidade

USP/FM/DBD-437/22

Responsável: Erinalva da Conceição Batista, CRB-8 6755

Dedicatória

À minha avó, Dorothy Barbieri (*in memorian*), minha fonte de inspiração.

À minha tia, Profa. Dra. Silvia Rodrigues Machado, amálgama de tia, mãe e referência profissional.

Ao meu querido filho, Antonio, luz em minha vida.

Agradecimentos

À Universidade pública, gratuita e de qualidade deste país.

À Universidade de São Paulo (USP) e ao Programa de Pós-Graduação em Patologia pela oportunidade de realização do curso de Doutorado.

À Profa. Dra. Vera Luiza Capelozzi a quem tenho uma enorme dívida de gratidão pela generosidade e pelo inestimável apoio e comprometimento científico.

À equipe do Laboratório de Patologia e Genômica Pulmonar pela amizade e pelo trabalho profissional e diligente, especialmente, às queridas amigas Tabatha Prieto e Camila Baldavira.

Ao Prof. Dr. Alexandre Todorovic Fabro pelos conselhos, incentivo e contribuição científica contínuos.

À Profa. Dra. Cláudia Aparecida Rainho pelo apoio científico e por colocar à disposição o Laboratório de Epigenética (Instituto de Biociências-UNESP-Botucatu) para o desenvolvimento deste estudo.

Ao Hospital das Clínicas da Faculdade de Medicina de Botucatu (HCFMB-UNESP), especialmente, ao Prof. Dr. André Luís Balbi, Superintendente do HCFMB, e à equipe do Departamento de Gestão de Atividades Acadêmicas (DGAA) pelo apoio ao meu desenvolvimento profissional.

Ao Laboratório de Consultoria em Patologia Bacchi, especialmente ao Prof. Dr. Carlos Eduardo Bacchi e à Profa. Dra. Maura Moscardi Bacchi pelos meus primeiros passos na Patologia Molecular, e às mentoras Dra. Lucimara Chioato Benine, Dra. Luciana Hayashi e Dra. Elida Paula Benquique Ojopi pelos ensinamentos nas técnicas de biologia molecular.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (2018/20403-6) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (483005/2012-6) pelo apoio financeiro para a realização desta pesquisa.

Finalmente, gostaria de agradecer à minha família pelo extraordinário incentivo e, principalmente, à minha tia, Silvia Rodrigues Machado e ao meu marido, Antonio Rugolo Junior, pelo incondicional apoio nesta minha trajetória.

NORMALIZAÇÃO ADOTADA

Esta tese está de acordo com as seguintes normas, em vigor no momento desta publicação:

Universidade de São Paulo. Faculdade de Medicina. Divisão de Biblioteca e Documentação. *Guia de apresentação de dissertações, teses e monografias.*
Elaborado por Anneliese Carneiro da Cunha, Maria Julia de A. L. Freddi, Maria F. Crestana, Marinalva de Souza Aragão, Suely Campos Cardoso, Valéria Vilhena. 3a ed. São Paulo: Divisão de Biblioteca e Documentação; 2011.

Referências: adaptado de *International Committee of Medical Journals Editors* (Vancouver).

Abreviaturas dos títulos dos periódicos de acordo com *List of Journals Indexed in Index Medicus.*

SUMÁRIO

Lista de Siglas e Abreviaturas

3′UTR - 3’região não traduzida

5'UTR - 5’região não traduzida

ACMG - Colégio Americano de Genética Médica e Genômica

ADC - Adenocarcinoma

ANVISA - Agência Nacional de Vigilância Sanitária

CEC - Carcinoma de células escamosas

CGC - Carcinoma de grandes células

CHT - Quimioterapia

CONITEC - Comissão Nacional de Incorporação de Tecnologias no SUS

COSMIC - Catálogo de Mutações Somáticas em Câncer

CPC - Carcinoma de pequenas células

CPCNP - Câncer de pulmão de células não pequenas

DNA - Ácido Desoxirribonucleico

EGF - Fator de crescimento epidérmico

EGFR- Receptor do fator de crescimento epidérmico

EGFR-TKIs- Inibidores da tirosina quinase associada a EGFR

EMT- Transição Epitélio-Mesenquimal

FDA - do inglês, “*Food and Drug Administration*” (Administração de Alimentos e Medicamentos)

GERP - do inglês, “*Genomic Evolutionary Rate Profiling*” (Perfil de Taxa Evolucionária Genômica)

HR - do inglês, “*Hazard Ratio*” (Razão de Risco)

HWE- do inglês, “*Hardy–Weinberg equilibrium*” (Equilíbrio de Hardy-Weinberg)

IC - Intervalo de confiança

ICI - Inibidores de pontos de controle imunológicos

INCA - Instituto Nacional de Câncer José Alencar Gomes da Silva

INDEL - Mutações de inserção ou deleção

miRNAs - microRNAs

NGS- do inglês, “*Next Generation Sequencing*” (Sequenciamento de Próxima Geração ou Sequenciamento de Nova Geração)

PD-L1 - Ligante de morte programada 1

PPS - Sobrevida pós-progressão

SG - Sobrevida global

SNP - Polimorfismos de Nucleotídeo Único

SNV - Variantes de Nucleotídeo Único

TCGA - do inglês, "*The Cancer Genome Atlas*" (Atlas do Genoma do Câncer)

TGFα - Fator de Transformação do Crescimento α

TMA - do inglês, "*Tissue microarray*" (microarranjos de tecido)

TNM - Sistema Tumor-Nódulo-Metástase

VUS - Variante de Significado Desconhecido

Lista de Quadros e Tabelas

Lista de Figuras

Resumo

Rugolo JM. *Perfil molecular das vias de sinalização dos carcinomas de pulmão de células não pequenas por Sequenciamento de Nova Geração (NGS) com ênfase à Transição Epitélio Mesenquimal (EMT)* [tese]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2022.

O câncer de pulmão (CP) é um dos tumores malignos mais comuns e uma das principais causas de morte relacionada ao câncer no mundo. O CP é classificado histologicamente em dois grupos: Carcinoma Pulmonar de Pequenas Células (CPPC) e Carcinoma Pulmonar de Células não Pequenas (CPCNP), sendo este o responsável por 85% dos carcinomas primários de pulmão. Um dos principais desafios do CP é o seu diagnóstico tardio, no qual a doença se encontra em estágio avançado impedindo o tratamento curativo. Uma melhor compreensão da patogênese da doença é necessária para auxiliar o diagnóstico, a seleção do melhor tratamento e o desenvolvimento de novas modalidades terapêuticas, a fim de melhorar os resultados do paciente. Com o advento da tecnologia do Sequenciamento de Nova Geração (NGS) uma compreensão abrangente da patogênese molecular da doença foi alcançada e novos biomarcadores emergiram para auxiliar no manejo da doença, oferecendo terapias eficazes e menos tóxicas. O objetivo geral deste trabalho foi avaliar o perfil de alteração genética em CPCNP utilizando a técnica de NGS, principalmente, a alteração de genes envolvidos no processo biológico conhecido como Transição Epitélio Mesenquimal (EMT)**.** Os objetivos específicos foram: Traçar o perfil imunológico do CPCNP utilizando a imunofluorescência multiplex e correlacionar os achados com o perfil de alteração genética relacionado à EMT; Investigar a influência de variantes não-codificantes do gene *PD-L1* em amostras de CPCNP e correlacionar com expressão proteica e as características clinicopatológicas; Estimar a frequência de mutação no gene *EGFR*, com destaque para as mutações de significado clínico incerto no CPCNP. A extração de DNA total de 70 amostras de CPCNP foi realizada a partir do kit QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). A qualidade das amostras foi avaliada utilizando Qubit® 3.0 Fluorometer (Invitrogen, Life Technologies, CA, USA). O sequenciamento foi realizado através da plataforma de NGS Illumina MiSeq V2 (2x 150, paired-end) e as bibliotecas genômicas foram construídas a partir do TruSeq Custom Amplicon v1.5 kit (TSCAP, Illumina, SanDiego, CA). Variantes genéticas com frequência populacional maior que 1% (popfreq_all >0,01) foram consideradas variantes germinativas e menor que 1% foram classificadas como variantes somáticas. As variantes encontradas foram comparadas com resultados depositados em bancos de dados públicos. Os resultados obtidos para cada objetivo específico foram compilados em três artigos. *Artigo 1)* Foram encontradas frequências semelhantes de mutações nos genes codificadores de proteínas de checkpoint imunológico e genes relacionados à EMT em nossa coorte e na coorte do TCGA. Os genes mais frequentemente mutados em nossa coorte foram *ZEB2* (n = 10; 14%), *MMP2* (n = 4; 6%), *ZEB1* (n = 2; 3%), *CDH1* (n = 2; 3%) e *CD44* (n = 2; 3%), todos envolvidos em EMT. Além disso, a variante *CTLA4* rs231775 com genótipo AA foi associada a menor sobrevida global (SG), quando comparado ao alelo G (HR = 2,091, IC de 95% = 0,233-2,219, P = 0,05). *Artigo 2)* As variantes de PD-L1 estudadas não interferiram na expressão da proteína, no entanto, as variantes rs4742098A>G, rs4143815G>C e rs7041009G>A foram associadas à recidiva da doença (P=0,01; P=0,05; P=0,02, respectivamente). No modelo de regressão de Cox, o genótipo GG da variante rs7041009 influenciou a SG (P<0,01), atuando como um fator co-dependente associado à radioterapia e recidiva em pacientes com CPCNP. *Artigo 3)* A taxa geral de mutação foi de 32,9%, dos quais 20,0% dos casos abrigavam mutações clássicas envolvendo

os exons 18-21 do *EGFR*. Mutações abrangendo outros exons foram identificadas em 12,9% dos casos. 11 variantes foram classificadas como de significado clínico desconhecido. Destas, cinco foram previstas como patogênicas de acordo com análises *in silico*. Conclui-se que este trabalho contribui na compreensão da patogênese do câncer de pulmão e destaca a importância da utilização do NGS e ferramentas computacionais no manejo clínico de pacientes portadores de CPCNP.

Descritores: Carcinoma pulmonar de células não pequenas; Sequenciamento de nucleotídeos em larga escala; Biologia computacional; Medicina de precisão; Biomarcadores; Mortalidade.

Abstract

Rugolo JM. *Molecular profile of signaling pathways of non-small cell lung carcinomas by Next Generation Sequencing (NGS) with Emphasis on Mesenchymal Epithelium Transition (EMT)* [thesis]. São Paulo: "Faculdade de Medicina, Universidade de São Paulo"; 2022.

Lung cancer (LC) is one of the most common malignant tumors and a leading cause of worldwide cancer-related death. LC is histologically classified into two groups: Small Cell Lung Carcinoma (SCLC) and Non-Small Cell Lung Carcinoma (NSCLC), responsible for 85% of primary lung carcinomas. One of the main challenges of LC is its late diagnosis, in which the disease is in an advanced stage, preventing curative treatment. A better understanding of disease pathogenesis is needed to aid diagnosis, selection of the best medicine, and development of new therapeutic modalities to improve patient outcomes. With the advent of Next Generation Sequencing (NGS) technology, a comprehensive understanding of the molecular pathogenesis of the disease has been achieved, and new biomarkers have emerged to aid in the management of the disease, offering effective and less toxic therapies. The general objective of this work was to evaluate the profile of genetic alteration in NSCLC using the NGS technique, mainly the alteration of genes involved in the biological process known as Epithelium Mesenchymal Transition (EMT). The specific objectives were: To trace the immunological profile of NSCLC using multiplex immunofluorescence and to correlate the findings with the profile of genetic alteration related to EMT; To investigate the influence of non-coding variants of the *PD-L1* gene on NSCLC samples and correlate with protein expression and clinicopathological features; To estimate the frequency of mutation in the *EGFR* gene, with emphasis on mutations of uncertain clinical significance in NSCLC. Total DNA extraction from 70 NSCLC samples was performed using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). Sample quality was assessed using Qubit® 3.0 Fluorometer (Invitrogen, Life Technologies, CA, USA). Sequencing was performed using the Illumina MiSeq V2 NGS platform (2x 150, paired-end), and the genomic libraries were built using the TruSeq Custom Amplicon v1.5 kit (TSCAP, Illumina, SanDiego, CA). Genetic variants with a population frequency greater than 1% (popfreq_all >0.01) were considered germline variants, and less than 1% were classified as somatic variants. The variants found were compared with results deposited in public databases. The results obtained for each specific objective were compiled in three articles. *Article 1)* Similar frequencies of mutations were found in genes encoding immunological checkpoint proteins and EMT-related genes in our and TCGA cohorts. The most frequently mutated genes in our cohort were *ZEB2* (n = 10; 14%), *MMP2* (n = 4; 6%), *ZEB1* (n = 2; 3%), *CDH1* (n = 2; 3%) and *CD44* (n = 2; 3%), all involved in TMS. In addition, the *CTLA4* rs231775 variant with the AA genotype was associated with lower overall survival (OS) compared to the G allele (HR = 2.091, 95% CI = 0.233-2.219, P = 0.05). *Article 2)* The *PD-L1* variants studied did not interfere with protein expression; however, the rs4742098A>G, rs4143815G>C, and rs7041009G>A variants were associated with disease recurrence (P=0.01; P=0. 05; P=0.02, respectively). In the Cox regression model, the GG genotype of the rs7041009 variant influenced OS (P<0.01), acting as a codependent factor associated with radiotherapy and relapse in NSCLC patients. *Article 3)* The overall mutation rate was 32.9%, of which 20.0% of the cases harbored classic mutations involving exons 18-21 of the *EGFR*. Mutations spanning other exons were identified in 12.9% of cases. Eleven variants were classified as of unknown clinical significance. Five were predicted to be pathogenic according to in silico analyses. It is concluded that this work contributes to understanding lung cancer's pathogenesis and highlights the importance of using NGS and computational tools in the clinical management of patients with NSCLC.

Descriptors: Carcinoma, non-small-cell lung; High-throughput nucleotide sequencing; Computational biology; Precision medicine; Biomarkers; Mortality.

Apresentação

Esta tese é composta pela compilação de três artigos a seguir que serão descritos e analisados: 1. Variants in Epithelial-Mesenchymal Transition and Immune Checkpoint Genes Are Associated with Immune Cell Profiles and Predict Survival in Non–Small Cell Lung Cancer; 2. Relevance of *PD-L1* Non-Coding Polymorphisms on the Prognosis of a Genetically Admixed NSCLC Cohort; 3. Clinical Outcome of Brazilian Patients with Non-Small Cell Lung Cancer in Early Stage Harboring Rare Epidermal Growth Factor Receptor Mutations.

INTRODUÇÃO

O câncer de pulmão é um dos tumores malignos mais comuns e uma das principais causas de morte relacionada ao câncer em homens e mulheres em todo o mundo, com crescimento no número de casos desde 2011 (1). No Brasil, segundo estimativas do Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA), o câncer de pulmão é a segundo mais comum na população, estimando-se, para o ano de 2020, 30.200 novos casos (2).

Essa neoplasia é uma patologia heterogênea e pode ser classificada histologicamente em dois grupos majoritários: Carcinoma de Pequenas Células (CPC) e Carcinoma de Células não Pequenas (CPCNP). O CPCNP é responsável por 85% dos carcinomas primários de pulmão e produz a maior taxa de mortalidade em todo o mundo (3). Dentre o grupo de CPCNP, podemos encontrar os seguintes grupos sub-histológicos: adenocarcinomas (ADC), carcinoma de células escamosas (CEC) e carcinoma de grandes células (CGC) (3).

O surgimento da doença pode estar associado a exposição de produtos químicos (por exemplo, carcinógenos associados ao tabaco e benzeno), metais pesados (como o rádio, cádmio e arsênio), radiações, agentes microbianos naturais, como determinadas aflotoxinas e vírus (Sarcoma de Rous, Hepatite, Papiloma humano). Todos estes agentes podem levar a uma má incorporação de base no DNA levando a mutações, que finalmente podem resultar em carcinomas pulmonares (4, 5). Porém, de todos os agentes, o câncer de pulmão está principalmente associado à exposição ao tabagismo; no entanto, tem sido observado um aumento da incidência de casos em não fumantes nas últimas décadas (5).

Atualmente, um dos principais desafios que vem sendo enfrentado no contexto do câncer de pulmão é o diagnóstico tardio e a doença avançada na apresentação assintomática, refletindo em um desfecho clínico com uma sobrevivência total de 5 anos de 16-20% (6, 7). Assim, uma melhor compreensão da patogênese do câncer de pulmão faz-se necessária para auxiliar o

diagnóstico oportuno, a seleção do tratamento do câncer e o desenvolvimento de novas modalidades terapêuticas, a fim de melhorar os resultados do paciente (8).

A busca pela base genômica do câncer tem se expandido exponencialmente desde a introdução das técnicas de sequenciamento de DNA no final da década de 1970 (8), e atualmente, com o uso de tecnologias de alto rendimento como o sequenciamento paralelo em larga escala ou *Next Generation Sequencing* (NGS), iniciou-se uma nova etapa no mapeamento das alterações genéticas responsáveis pelo processo da carcinogênese (9). Essa tecnologia, disponível desde 2004 (9, 10), revolucionou o campo da pesquisa genômica na era da Medicina de Precisão, permitindo sequenciar o genoma completo de tumores individuais na busca de mutações tornando-se uma ferramenta fundamental no diagnóstico precoce e na escolha da terapêutica apropriada, também conhecida como terapias-alvo (10).

Em câncer de pulmão, os principais genes conhecidos com potenciais mutações são *EGFR*, *KRAS*, *FGFR1*, *ERBB2*, *PIK3CA*, *ALK*, *BRAF*, *ROS1*, *MAP2K1*, *RET*, *NRAS* e *AKT1* (10-12). A descoberta da ativação do receptor do fator de crescimento epidérmico (EGFR) via mutação, amplificação e ou sua superexpressão tem sido relatada e implicada na patogênese de muitas malignidades humanas, incluindo CPCNP (13). Alguns estudos demonstraram que a expressão de *EGFR* no CPCNP está associada à redução da sobrevida, metástase e sensibilidade deficiente à quimioterapia (14).

Estudos recentes identificaram que a mutações em *EGFR* podem ser responsáveis pela resistência adquirida aos inibidores tirosino quinase (EGFR-TKI), amplamente utilizados atualmente no tratamento do CPCNP. Representando cerca de 50% dos casos de resistência adquirida pelos inibidores de *EGFR* (15).

Outro ponto relevante que vem sendo observado no câncer de pulmão são as moléculas de ponto de controle imunológico envolvidas na evasão imune de tumor e os inibidores de ponto de controle imunológico (ICIs), introduzidos recentemente na imunoterapia antitumoral. Nessa nova abordagem, tem-se como alvo um receptor inibitório, o receptor de morte celular programada-1 (PD-

1), para auxiliar o sistema imunológico na identificação e neutralização de células malignas. No entanto, as células tumorais podem escapar da vigilância imunológica do hospedeiro expressando o ligante 1 de morte programado 1 (PD-L1) como um mecanismo resistente adaptável para suprimir este receptor inibitório (16). Contudo, ainda são vários os desafios necessários para suportar o uso deste componente como biomarcador.

Assim, diante do exposto e devida a elevada incidência e mortalidade do CPCNP, vemos a necessidade de contribuir com novos estudos translacionais com o emprego de tecnologias de alto rendimento, a fim de buscar uma melhor compreensão da natureza molecular deste câncer, destacar genes alterados previamente não apreciados, e investigar os possíveis determinantes genéticos envolvidos no desenvolvimento de resistência a terapias, contribuir para que no futuro se encontre novos biomarcadores preditivos de resposta terapêutica e novas abordagens terapêuticas para prevenir ou pelo menos atrasar o aparecimento de resistências e regressão tumoral.

REVISÃO DA LITERATURA

2.1 Epidemiologia do câncer de pulmão

O câncer de pulmão é a principal causa de morte por câncer em homens e mulheres em todo o mundo (1, 17). Em 2020, em homens, o câncer de pulmão foi a principal causa de morte por câncer em 93 países, seguido pelo câncer de próstata (48 países) e câncer de fígado (23 países). Já em mulheres, o câncer de mama (em 110 países) e o câncer do colo do útero (em 36 países) foram as principais causas de morte por câncer, o câncer de pulmão ocupando o terceiro lugar, em 25 países (18), como ilustrado na **Figura 1.**

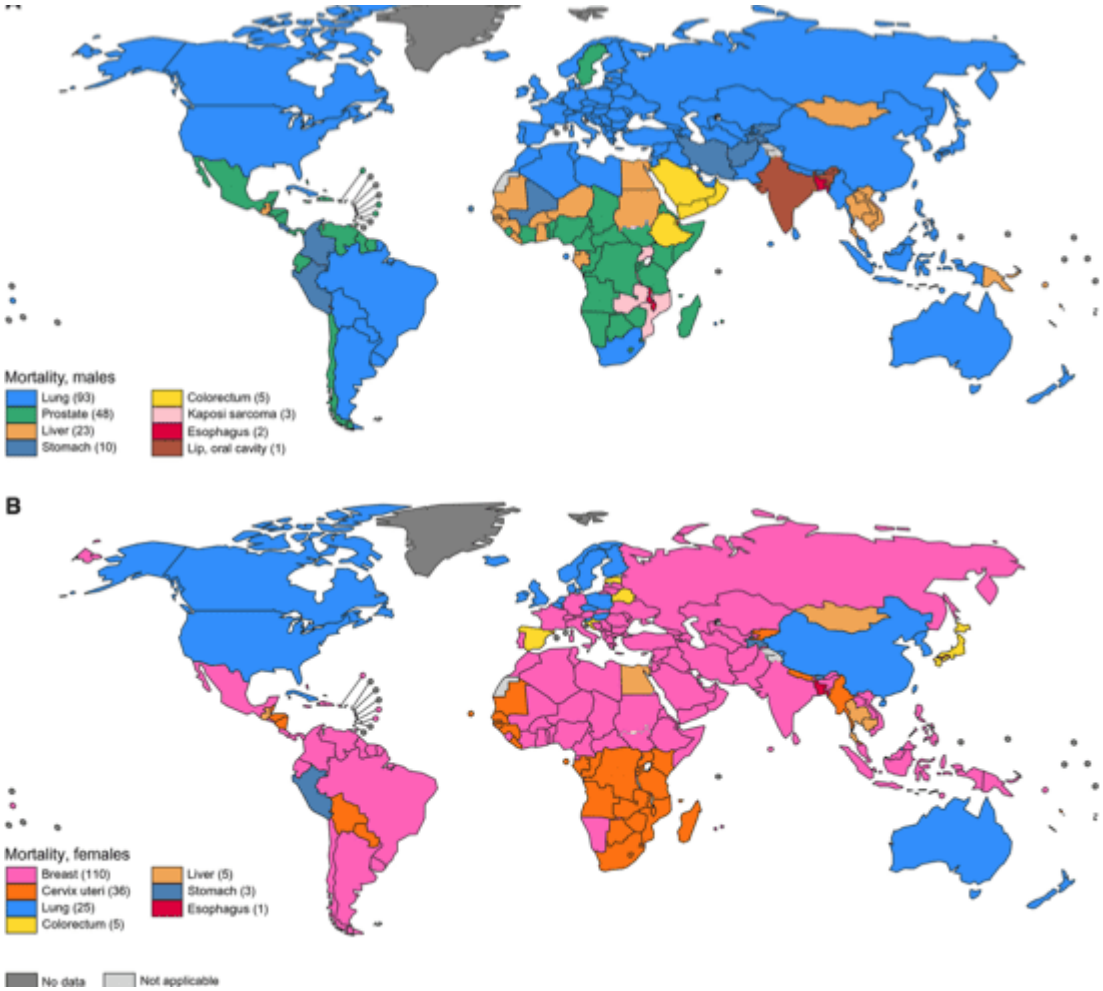


Figura 1: Mortalidade por câncer no mundo em 2020 entre (A) homens e (B) mulheres. Número de países representados em cada tipo de câncer estão incluídos na legenda.

Fonte: Adaptado de Sung et al. (18).

Nos Estados Unidos, aproximadamente um quarto (22%) de todas as mortes por câncer são devido ao câncer de pulmão tanto em homens quanto em mulheres, conforme mostrado na Error! Reference source not found. (19).

Estimated Deaths

Males				Females		
Lung & bronchus	69,410	22%		Lung & bronchus	62,470	22%
Prostate	34,130	11%		Breast	43,600	15%
Colon & rectum	28,520	9%		Colon & rectum	24,460	8%
Pancreas	25,270	8%		Pancreas	22,950	8%
Liver & intrahepatic bile duct	20,300	6%		Ovary	22,950	5%
Leukemia	13,900	4%		Uterine corpus	12,940	4%
Esophagus	12,410	4%		Liver & intrahepatic bile duct	9,930	3%
Urinary bladder	12,260	4%		Leukemia	9,760	3%
Non-Hodgkin lymphoma	12,170	4%		Non-Hodgkin lymphoma	8,550	3%
Brain & other nervous system	10,500	3%		Brain & other nervous system	8,100	3%
All Sites	**319,420**	**100%**		**All Sites**	**289,150**	**100%**

Figura 2: Os 10 principais tipos de câncer para casos de morte por sexo nos Estados Unidos, 2021. As estimativas são arredondadas para o 10 mais próximo. A classificação é baseada em projeções modeladas e pode diferir dos dados mais recentes observados.

Fonte: Adaptado de Siegel et al. (19).

Já no Brasil, com relação à mortalidade, em 2019, o câncer de pulmão foi a principal causa de morte por câncer em homens (13,8%), correspondendo a 16.733 casos; e o segundo em mulheres (11,4%; 12.621 casos), após o câncer de mama com 18.068 casos (16,4%). Para o triênio 2020-2022, estimam-se, para cada ano do triênio, 17.760 novos casos de câncer de pulmão em homens e 12.440 em mulheres. Atualmente, no país, o câncer de pulmão ocupa a terceira posição de câncer mais incidente em homens e o quarto lugar para as mulheres, como ilustrado pela **Error! Reference source not found.** (20).

Localização Primária	Casos	%			Localização Primária	Casos	%
Próstata	65.840	29,2%	Homens	Mulheres	Mama feminina	66.280	29,7%
Cólon e reto	20.520	9,1%			Cólon e reto	20.470	9,2%
Traqueia, brônquio e pulmão	17.760	7,9%			Colo do útero	16.590	7,4%
Estômago	13.360	5,9%			Traqueia, brônquio e pulmão	12.440	5,6%
Cavidade oral	11.180	5,0%			Glândula tireoide	11.950	5,4%
Esôfago	8.690	3,9%			Estômago	7.870	3,5%
Bexiga	7.590	3,4%			Ovário	6.650	3,0%
Linfoma não Hodgkin	6.580	2,9%			Corpo do útero	6.540	2,9%
Laringe	6.470	2,9%			Linfoma não Hodgkin	5.450	2,4%
Leucemias	5.920	2,6%			Sistema nervoso central	5.220	2,3%

Figura 3: Distribuição proporcional dos dez tipos de câncer mais incidentes estimados para 2020 por sexo, noBrasil.

Fonte: Adaptado de INCA (20).

2.2 Fatores de risco associados ao câncer de pulmão

O tabagismo é considerado o principal fator de risco para o desenvolvimento do câncer de pulmão (21), no qual quase 50% dos casos diagnosticados ocorrem em ex-fumantes (22).

O vício do tabagismo é mediado, principalmente, devido à presença de nicotina no tabaco e receptores nicotínicos de acetilcolina, que são encontrados não só em todo o sistema nervosos central, mas também em células não neurais, tais como células do tecido epitelial brônquico (23). No entanto, a nicotina em si não é responsável pelo desenvolvimento do câncer, a exposição ao alcatrão é que acarreta o carcinoma pulmonar (24).

O alcatrão é a matéria particulada da fumaça do cigarro, com cerca de 3500 compostos diferentes e a maioria deles são cancerígenos. Os principais carcinógenos no alcatrão podem ser amplamente classificados em hidrocarbonetos aromáticos policíclicos, nitrosaminas, aminas aromáticas, aminas aromáticas heterocíclicas e aldeídos. O alcatrão também inclui outros carcinógenos; compostos orgânicos como 1,3-butadieno e carbamato de etila; e compostos inorgânicos como arsênio, níquel, cromo, cádmio, hidrazina e polônio-210 (24-26). Substâncias que são capazes de ativar várias vias metabólicas e de sinalização e contribuir para a carcinogênese pulmonar. Dados sugerem que as nitrosaminas são responsáveis por reduzir drasticamente o nível da proteína 10kDa em células de Clara, que são células excretoras, encontradas no epitélio dos bronquíolos primários, responsáveis por excretar e absorver glicoproteínas que revestem esse epitélio.

A proteína 10kDa é considerada uma proteína multifuncional, secretada naturalmente pelo epitélio das vias aéreas de mamíferos e a perda de sua expressão por parte das células de Clara provoca uma incidência maior de hiperplasia epitelial e de adenoma de vias aéreas (23). As nitrosaminas também atuam como agonistas beta-adrenérgicos, que podem ativar a via de sinalização do receptor do fator de crescimento epidérmico (EGFR), resultando na ativação exacerbada das vias de sobrevivência e proliferação celular (23).

Apesar da compreensão do câncer de pulmão estar historicamente ligada ao tabagismo, no entanto, nas últimas décadas, tem sido observado o aumento da incidência de casos em não fumantes (5). Aproximadamente 10% de todos os pacientes com câncer de pulmão negam o tabagismo. Na Ásia, por exemplo, a proporção de pacientes não tabagistas é de 30 a 40%. Estudos epidemiológicos mostram uma maior proporção de mulheres no grupo de pacientes não tabagistas acometidos por câncer de pulmão em escala mundial (23).

Na ausência do histórico de tabagismo, outros fatores de risco podem estar associados ao câncer de pulmão. Estudos mostram que 25% dos casos de câncer de pulmão em pacientes não tabagistas estão associados ao tabagismo passivo (23), além de outros fatores associados, como a exposição ao radônio e a fumaça de combustível doméstico (21). As mutações e alguns polimorfismos de nucleotídeo único (SNP; do inglês *single-nucleotide polymorphisms*) também aumentam o risco de câncer de pulmão (24). Pacientes portadores da Síndrome de Peutz-Jeghers, doença hereditária autossômica dominante, que é caracterizada pelo surgimento precoce de pólipos no intestinais, têm 15% de chance de desenvolver câncer de pulmão ao longo da vida. Esta doença está relacionada a mutações presentes no gene *LKB1*, um supressor tumoral responsável por codificar uma proteína com atividade de serina/treonina quinase, envolvida em várias vias de sinalização, incluindo a mTOR (23).

Além disso, infecções como Mycobacterium tuberculosis e o Vírus do Papiloma Humano (HPV) também são fatores de risco. Outras associações menos conclusivas incluem doenças inflamatórias, como asma e sarcoidose (21).

2.3 Classificação anatomopatológica do câncer de pulmão

O câncer de pulmão é classificado de acordo com o seu tipo histológico. O câncer de pulmão compreende vários tipos histológicos, porém os carcinomas representam o principal tipo histológico da neoplasia pulmonar, considerando-se que os demais tipos existentes compreendem menos de 1% de todos os casos.

Há dois principais grandes tipos de carcinomas pulmonares: Carcinoma Pulmonar de Pequenas Células (CPPC) e Carcinoma Pulmonar de Células não Pequenas (CPCNP), sendo este último, o responsável por 85% dos casos de câncer de pulmão (27, 28).

O câncer de pulmão pode surgir a partir dos grandes brônquios (tumores centrais) ou pequenos brônquios, bronquíolos e alvéolos (tumores periféricos) (23). O CPPC é originário, principalmente, das vias respiratórias centrais e derivado de células neuroendócrinas (29).

Dentre o grupo dos CPCNP, podemos encontrar vários subgrupos histológicos, sendo os mais comuns: Adenocarcinoma (ADC), Carcinoma de Células Escamosas ou Carcinoma Espinocelular (CEC) e Carcinoma Indiferenciado de Grandes Células (CGC) (27, 30). Dentre eles, o ADC é o mais comum, sendo responsável por aproximadamente 40% de todos os casos pulmonares (31). Normalmente, este tipo histológico se origina do epitélio glandular, de células-tronco broncoalveolares, células de Clara ou pneumócitos tipo II na periferia pulmonar. O CEC se desenvolve de células epiteliais brônquicas, principalmente nas vias aéreas centrais e brônquios segmentares, sendo responsável por aproximadamente 20% de todos os casos de câncer de pulmão. (28, 32). Já o CGC é um tumor epitelial maligno que compreende grandes células poligonais mostrando nenhuma evidência óbvia de diferenciação histológica. O carcinoma de células grandes é caracterizado por grupos dispersos de grandes células não diferenciadas, polimórficas e frequentemente de núcleo duplo ou múltiplo (33). Ainda no grupo do CPCNP, existe um predomínio de CEC entre pacientes tabagistas, com uma frequência de 53%, comparado a 19% de ADC. Em relação aos não tabagistas, 62% dos casos compreendem o ADC, contra 18% de CEC (23). Já em relação ao sexo, o tipo histológico CEC é o mais

frequente em homens com 44%, contra 25% em mulheres. Esta relação está invertida no ADC, com 42% dos casos em mulheres, contra apenas 28% em homens.

Contudo, é fato que a epidemiologia do câncer de pulmão vem sofrendo mudanças. Nos anos de 1950, era uma doença predominantemente de homens fumantes com CEC. Atualmente, é uma doença que inclui mulheres, não fumantes e com maior incidência de ADC (23).

2.4 Estadiamento e prognóstico do câncer de pulmão

O estadiamento do tumor é o fator prognóstico mais importante no câncer de pulmão. A Associação Internacional para o Estudo do Câncer de Pulmão (*International Association for the Study of Lung Cancer*, IASLC) recomenda que o estadiamento do câncer de pulmão seja realizado seguindo a classificação do sistema de estadiamento tumor-nódulo-metástase (TNM) da *American Joint Committee on Cancer* (AJCC), dadas as grandes diferenças na sobrevida em relação ao estágio do tumor (34, 35).

O sistema TNM baseia-se no tamanho, localização e extensão regional do tumor primário; localização dos linfonodos malignos regionais que drenam a região; e ausência/presença de metástases distantes conforme mostrado no

Quadro 1. Atualmente, o sistema se encontra em sua oitava edição para câncer de pulmão, a qual foi publicada no final de 2016 e promulgada em janeiro de 2017, substituindo as versões anteriores (36).

Um dos principais desafios que se enfrenta no câncer de pulmão é que a maioria dos casos se encontra em estágio avançado ao diagnóstico, refletindo em uma taxa de sobrevida de 5 anos de 10 a 20%. No Brasil, essa taxa é de 18%, se mostrando em consonância com as taxas mundiais (7, 37).

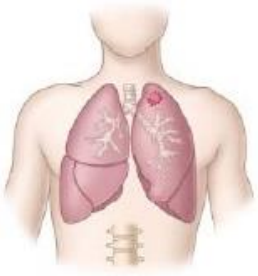
Com a doença estabelecida, os pacientes podem ser divididos em três grupos que, de maneira diferenciada, refletem a extensão da doença e abordagem terapêutica.

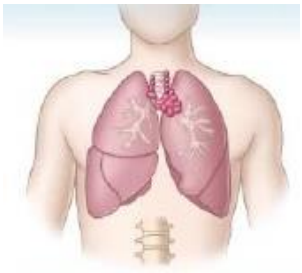
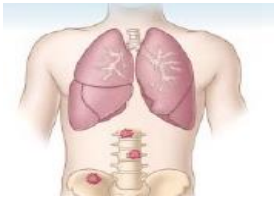
O primeiro grupo compreende aqueles tumores cirurgicamente ressecáveis, em geral no estágio I e II. Este é o grupo de melhor prognóstico com taxas de sobrevida em cinco anos de 67% para o estágio I, 37% para o estágio IIa e 24% para o estágio IIb. Nos estágios I e II, o prognóstico é influenciado adversamente pela presença de sintomas pulmonares, tamanho do tumor maior que 3 centímetros, acometimento linfonodal e sexo masculino com idade maior que 60 anos (38).

O segundo grupo de pacientes inclui tumores avançados local ou regionalmente, compreendendo o estágio III, o qual subdivide-se em IIIa e IIIb. O estágio IIIa é caracterizado pela presença de linfonodo mediastinal ipsilateral acometido (N2) ou por lesão T3 com linfonodo hilar positivo (N1). Já o estágio IIIB é caracterizado pela presença de envolvimento de linfonodo mediastinal contralateral ou supraclavicular (N3) ou tumor T4. A taxa de sobrevida global em cinco anos para pacientes IIIa varia de 10 a 15%, já o paciente em estágio IIIb, isto é, com acometimento de linfonodos mediastinais, é de 2 a 5% (38).

Finalmente, o último grupo de pacientes, compreende aqueles pacientes que apresentam metástases (M1) ao diagnóstico. Neste estágio, o prognóstico é ruim, principalmente, em pacientes com baixo performance status (KPS) e perda de peso maior que 10% (38).

Quadro 1. *Classificação do sistema de estadiamento tumor-nódulo-metástase (TNM) do câncer de pulmão.*

	T: **Tumor primário**
	**TX**: O tumor primário não pode ser avaliado ou o tumor originado pela presença de células malignas no escarro ou lavagens brônquicas, mas não pode ser visualizado por imagem ou broncoscopia **T0**: Sem evidência de tumor primário **Tis**: Carcinoma in situ **T1**: Tumor com maior dimensão ≤3 cm, rodeado por pleura visceral ou pulmonar, sem evidência broncoscópica de invasão mais proximal do que o brônquio lobar (ou seja, não no brônquio principal) **T1a (mi)**: Adenocarcinoma minimamente invasivo **T1a**: Tumor com maior dimensão ≤1 cm **T1b**: Tumor com maior dimensão >1 cm e ≤2 cm **T1c**: Tumor com maior dimensão >2 cm e ≤3 cm

	T2: Tumor> 3 cm e ≤5 cm com qualquer uma das seguintes características: # Envolvimento do brônquio principal, independentemente da distância da carina, mas sem envolvimento da carina; # Invasão da pleura visceral; # Associação com atelectasia ou pneumonite obstrutiva que se estende à região hilar, envolvendo parte ou todo o pulmão. **T2a:** Tumor com maior dimensão >3 cm e ≤4 cm **T2b:** Tumor com maior dimensão >4 cm e ≤5 cm **T3**: Tumor com maior dimensão >5 cm e ≤7 cm ou associado a nódulos tumorais separados no mesmo lobo que o tumor primário ou invada diretamente qualquer uma das seguintes estruturas: parede torácica (incluindo a pleura parietal e tumores do sulco superior), nervo frênico, pericárdio parietal **T4: Tumor com maior dimensão >7 cm ou associado a nódulos tumorais separados em um lobo ipsilateral diferente do tumor primário ou invada qualquer uma das seguintes estruturas: diafragma, mediastino, coração, grandes vasos, traquéia, nervo laríngeo recorrente, esôfago, corpo vertebral e carina**
	**N: Linfonodos regionais** **Nx**: Linfonodos regionais não podem ser avaliados **N0**: Ausência de metástases de linfonodo regional **N1**: Metástases em linfonodos ipsilaterais peribrônquicos e/ou ipsilaterais hilares e linfonodos intrapulmonares, incluindo envolvimento por extensão direta **N2**: Metástases em linfonodo (s) ipsilateral (is) mediastinal (is) e/ou subcarinal (is) **N3**: Metástases em linfonodo (s) ipsilateral (is) contralateral (is) mediastinal (is), contralateral (is) hilar (es), escaleno ipsilateral ou contralateral, ou linfonodo (s) supraclavicular (es)
	**M: Metástases distantes** **M0**: Ausência de metástases distantes **M1**: Presença de metástases distantes **M1a**: Nódulo (s) tumoral (is) separado (s) em um lobo contralateral; tumor com nódulo (s) pleural (is) ou derrame pleural (ou pericárdico) maligno **M1b**: Metástase extratorácica única **M1c: Múltiplas metástases extratorácicas em um ou mais órgãos**

Fonte: Adptado de Goldstraw et al. (36).

2.5 Alterações moleculares no câncer de pulmão

Nas últimas décadas, com o desenvolvimento das técnicas de sequenciamento genômico e de métodos de avaliação de alteração molecular, foi possível evidenciar que o câncer de pulmão, principalmente o CPCNP, é composto por subgrupos de alterações moleculares responsáveis por determinados fenótipos tumorais, influenciando, desta forma, o comportamento clínico destes tumores (39). Como exemplo clássico destas alterações, destacam-se as mutações condutoras (*drivers*) neste tipo de tumor (39-41).

Essas mutações ocorrem em genes que codificam proteínas de sinalização cruciais para a proliferação e sobrevivência celular; sendo necessárias para o surgimento do fenótipo maligno do tumor, influenciando fortemente nos resultados do tratamento do câncer (9). No câncer de pulmão, os principais genes conhecidos com potenciais mutações são: *TP53* (46%), *CDKN2A* (43%), *KRAS* (32%), *STK11* (17%), *EGFR* (11%), *MET* (7%), *BRAF* (7%), *PIK3CA* (4%), *PTEN* (3%), *ROS1* (1%), *ALK* (1%), *NRAS* (<1%) e *RET* (<1%) (42).

2.5.1 EGFR

O oncogene frequentemente mutado em carcinomas pulmonares de não tabagistas é o *EGFR*, com cerca de 50% dos casos. Na verdade, o *EGFR* é superexpresso em uma variedade de cânceres humanos, incluindo pulmão, cabeça e pescoço, cólon, pâncreas, mama, ovário, bexiga, rim e gliomas (43).

O *EGFR* pertence à família de receptores de tirosina quinases, também conhecida como família HER ou ErbB, representada por quatro membros conhecidos como: EGFR (HER1; ErbB1), HER2 (ErbB2), HER3 (ErbB3) e HER4 (ErbB4) (44).

A proteína EGFR consiste em três regiões e quatro domínios: uma região extracelular responsável pela ligação ao ligante, uma única região de hélice transmembrana e uma região citoplasmática. O domínio da tirosina quinase é responsável por aproximadamente 50% da região citoplasmática, com o restante composto pela região da justamembrana citoplasmática

(JM) de 38 aminoácidos e a região do terminal carboxila (CT) de 225 aminoácidos, conforme mostrado na **Figura 4** (45).

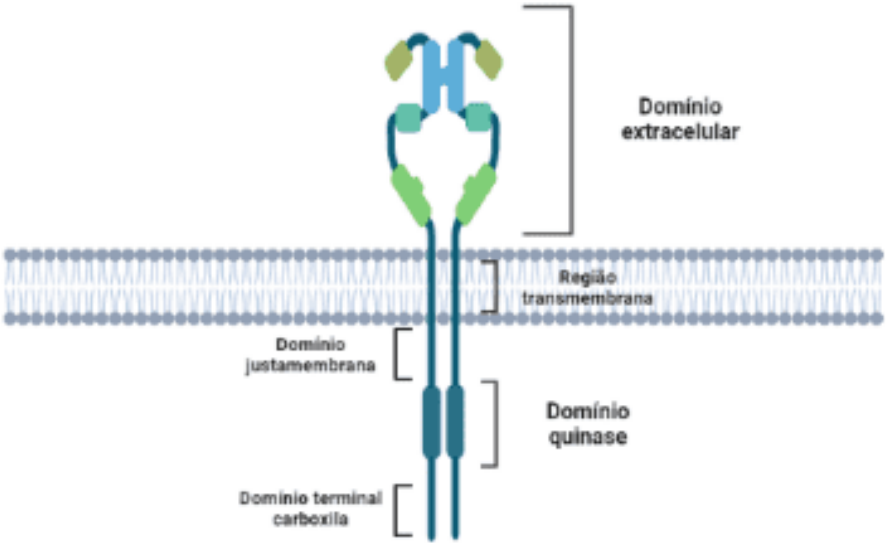


Figura 4: Regiões e domínios da proteína EGFR. Região extracelular (com o domínio extracelular), região de hélice transmembrana e região citoplasmática (com o domínio justamembrana, o domínio quinase e o domínio terminal carboxila).

Fonte: Próprio autor - criado em BioRender[1]

Como outros membros da família HER, o EGFR é um receptor transmembrana, ativado em resposta ao ligante, o qual se liga ao domínio extracelular do receptor (46). Alguns exemplos destes ligantes são: fatores de crescimento epidérmico (EGF), fatores de transformação de crescimento alfa (TGFα), fatores de crescimento semelhante ao EGF de ligação à heparina, anfirregulina, beta-celulina, epirregulina, epígeno e neuregulina 1-4 (47). O rearranjo conformacional extracelular após a ligação do ligante, permite a dimerização do receptor, levando os domínios da quinase citoplasmática a se transativar, conforme mostrado na **Error! Reference source not found.**

[1] https://biorender.com/

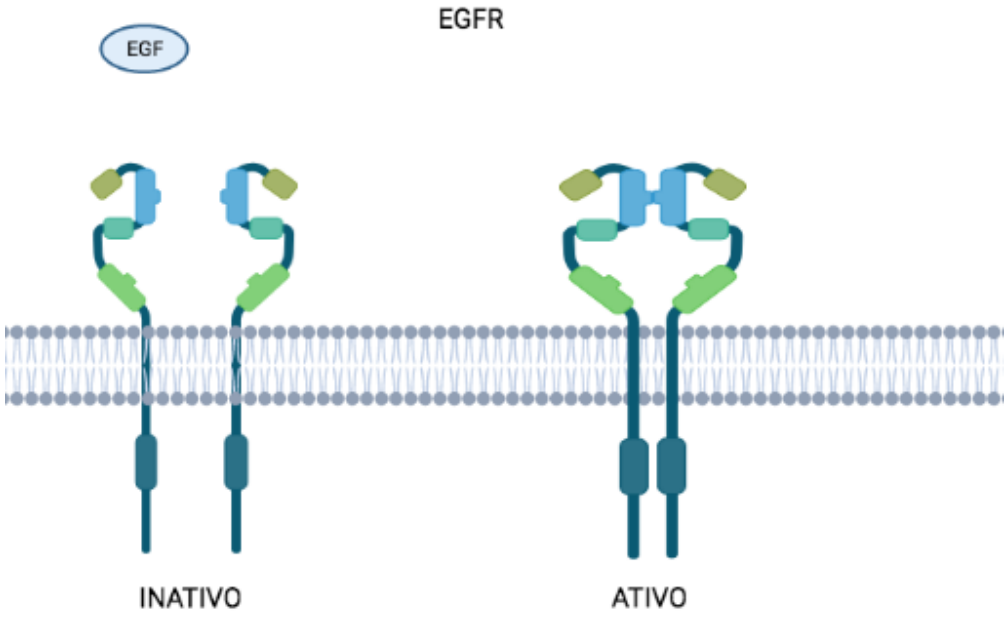


Figura 5: Diagrama esquemático da ativação de EGFR. São mostrados os domínios de EGFR: na região extracelular, a hélice transmembrana e a região justamembrana citoplasmática e o domínio da tirosina quinase. Na ausência de ligante, o EGFR reside na superfície da célula em uma conformação inativa/autoinibida (esquerda). Após a ligação do ligante (por exemplo, o EGF), a conformação autoinibitiva no ectodomínio EGFR é liberada, levando à dimerização do receptor mediada por ectodomínio (direita).

Fonte: Próprio autor - criado em BioRender

A ativação do domínio quinase do receptor leva à autofosforilação, a ativação de quinase e subsequente ativação das três principais vias de sinalização associadas ao EGFR, sendo a principal a via Ras/Raf/proteína quinase ativada por mitógeno (MAPK) e as PI3K/AKT e STATs, que conduzem o comportamento mitogênico, anti apoptótico, angiogênico e pró-invasivo das células cancerosas de acordo com a ilustração da **Error! Reference source not found.** (48).

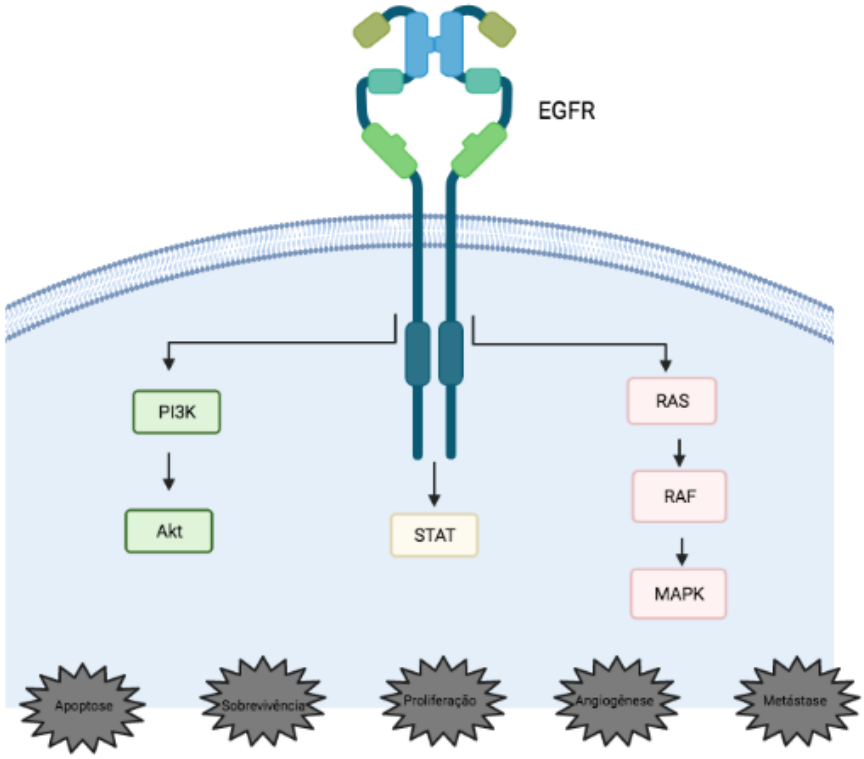


Figura 6: Principais mediadores à jusante da via de sinalização de EGFR no câncer de pulmão. É mostrado a ativação de quinase e a subsequente ativação das principais três principais vias de sinalização associadas ao EGFR: RAS/RAF/MEK/ERK1/2, PI3K/AKT e a via STAT.

Fonte: Próprio autor - criado em BioRender

No câncer de pulmão, a via EGFR pode se tornar constitutivamente ativa através de mutações neste gene. Tais mutações causam ativação da atividade de tirosina quinase do EGFR por desestabilizar sua conformação de auto inibição, que é normalmente mantida na ausência de estímulo por um ligante (23).

2.5.1.1 Mutações na região extracelular da proteína EGFR

Existem três tipos principais de mutações do tipo deleção na região extracelular, elas se diferem dependendo do local e da duração das deleções, sendo: EGFR vI, EGFR vII e EGFRvIII. Destas formas mutantes, o EGFRvIII é a mutação mais comum em gliomas (30-50%). Este mutante carece de uma grande parte da porção extracelular, incluindo a região de ligação ao ligante, levando à dimerização constitutiva e ativação do receptor. Esta mutação é detectada em 5% dos carcinomas de células escamosas de pulmão, mas não em outras histologias de células não pequenas. Além das deleções, novas mutações missense no domínio extracelular foram relatadas em 13,6% nos

glioblastomas; no entanto, essas mutações pontuais não foram encontradas em tumores de pulmão (49).

2.5.1.2 Mutações na região justamembrana da proteína EGFR

Há um domínio na região da justamembrana do EGFR que desempenha um papel ativador. Este domínio de ativação parece aumentar a formação do dímero assimétrico, promovendo assim a ativação do domínio quinase. Várias mutações raras neste domínio foram identificadas em pacientes com CPCNP, duas variantes missenses, V689M e L703F foram consideradas constitutivamente ativas (49).

2.5.1.3 Mutações no domínio quinase da proteína EGFR

Uma duplicação da quinase no domínio da tirosina quinase foi descrita em glioblastomas. Este mutante é constitutivamente ativo e confere tumorigenicidade. As mutações *EGFR* mais prevalentes em CPCNP, consistem em pequenas deleções *inframe* em torno do motivo LREA conservado do éxon 19 (resíduos 747-750) e uma mutação pontual (L858R) no éxon 21, que respondem por mais de 90% de todas as mutações da quinase EGFR. Mutações pontuais no éxon 18, predominantemente em G719, são responsáveis por aproximadamente outros 5% das mutações *EGFR*. As inserções *inframe* e as mutações pontuais no éxon 20 são responsáveis por 5% das mutações. Essas mutações de *EGFR* ativam a via de sinalização de EGFR promovendo sinais de pró-sobrevivência e antiapoptóticos mediados por EGFR por meio de alvos a jusante (49).

2.5.1.4 Mutações no domínio C-terminal intracelular da proteína EGFR

Em gliomas, foram relatadas duas formas de mutantes de deleção na região do terminal carboxila. O EGFR vIV, que contém uma deleção *inframe,* e o EGFR vV, que possui um truncamento do terminal carboxila. Essas variantes parecem ser constitutivamente ativas, já que

análises computacionais sugerem que isso se deve ao fato de que a região deletada tem um efeito inibitório sobre a atividade da quinase (49).

2.6 Aspectos terapêuticos no câncer de pulmão

O tratamento do câncer de pulmão é baseado no estágio da doença, o qual, por sua vez, é determinado pelo sistema de estadiamento TNM. Para as doenças de estágios I e II, a ressecção cirúrgica constitui o tratamento-padrão. No entanto, devido à alta taxa de pacientes diagnosticados tardiamente e com doença avançada, apenas uma pequena proporção de pacientes é submetida a cirurgia com intenção curativa (37). A doença de estágio III, que é inoperável, é tratada com quimioterapia e radioterapia. Já a doença no estágio IV é tratada com quimioterapia, radiação paliativa e terapia de suporte. E, embora o câncer de pulmão seja uma doença heterogênea, até meados dos anos 2000, o tratamento era baseado em quimioterapia citotóxica incluindo agentes alquilantes (cisplatina), antimetabólitos (gencitabina), inibidores de microtúbulos (paclitaxel), inibidores de topoisomerase (etoposídeo) e antibióticos citotóxicos (doxorrubicina) (10, 50). O regime quimioterápico ideal para o CPCNP ainda é tópico de investigação clínica.

No campo da medicina de precisão, surgiram os termos "terapias-alvo" ou “terapia dirigida” referindo-se às terapias de alvo molecular, nas quais os fármacos se ligam a proteínas de sinalização, interferindo, assim, no metabolismo, sobrevida e proliferação da célula. Neste contexto, com o conhecimento das mutações ativadoras do gene *EGFR*, alcançou-se notável progresso no tratamento do CPCNP com o advento dos inibidores de tirosina quinase (EGFR-TKIs, inibidores tirosina quinase), que são capazes de inibir a transdução de sinais da proteína EGFR. Dentre os pacientes com CPCNP, aqueles com tumores que apresentam mutações no gene *EGFR* podem se beneficiar do tratamento com EGFR-TKIs (51).

Os primeiros EGFR-TKIs foram sintetizados na década de 90. O gefitinibe (IRESSA®; AstraZeneca) foi o primeiro EGFR-TKIs aprovado pela Agência Federal do Departamento de Saúde e Serviços Humanos dos Estados Unidos (FDA, do inglês *Food and Drug Administration*) e, em

2003, foi aprovado para o tratamento de CPCNP avançado após o insucesso da terapia convencional (51). No Brasil, o gefitinibe foi aprovado pela Agência Nacional de Vigilância Sanitária (ANVISA) em 2011, e apenas em 2013, a Comissão Nacional de Incorporação de Tecnologias no SUS (CONITEC) recomendou a sua incorporação, em primeira linha para o tratamento do CPCNP avançado ou metastático com mutação *EGFR*, no entanto, sem a criação de um novo procedimento, sem a alteração dos valores dos procedimentos disponíveis e sem a modificação do modelo de financiamento da quimioterapia no SUS (52).

Em 2004, um outro EGFR-TKI, o erlotinibe (Tarceva®; Roche Produtos Químicos e Farmacêuticos SA) foi aprovado pelo FDA para o tratamento de CPCNP após a não resposta à quimioterapia citotóxica (51). No Brasil, sua aprovação ocorreu em 2006 para o tratamento de primeira linha de pacientes com CPCNP localmente avançado ou metastático, com mutações ativadoras de *EGFR*; terapia de manutenção a pacientes com CPCNP avançado ou metastático que não tenham progredido na primeira linha de quimioterapia; tratamento de pacientes com CPCNP localmente avançado ou metastático (estádios IIIb e IV), após a falha de pelo menos um esquema quimioterápico prévio. Mas, somente em 2013, a CONITEC recomendou a incorporação do erlotinibe, em primeira linha para o tratamento do CPCNP avançado ou metastático com mutação *EGFR*, no entanto, sem a criação de um novo procedimento, sem a alteração dos valores dos procedimentos disponíveis e sem a modificação do modelo de financiamento da quimioterapia no SUS (52). Tanto o gefitinibe quanto o erlotinibe tratam-se da primeira geração de EGFR-TKIs.

Já em 2014, o afatinibe (GIOTRIF®; Boehringer Ingelheim Pharmaceuticals), um bloqueador irreversível da família ErbB, também foi aprovado pelo FDA para uso clínico em pacientes que nunca foram submetidos a quimioterapia e cujos tumores abrigam mutações ativadoras do gene *EGFR* (51). As drogas bloqueadoras irreversíveis são consideradas EGFR-TKIs de segunda geração.

A inibição reversivelmente (pelo gefitinibe ou erlotinibe) ou irreversivelmente (pelo afatinibe) da atividade de tirosina quinase do EGFR ocorre pela competição destas drogas e moléculas de ATP por sítios de ligação no domínio C-terminal (sítios catalíticos) do receptor de EGFR. O bloqueio da fosforilação desses sítios impede a transdução de sinais através de componentes a jusante da via por meio do bloqueio da ativação das vias MAPK e PI3K, levando a redução da proliferação, sobrevivência e angiogênese de células cancerosas (51).

Atualmente, estamos na terceira geração de EGFR-TKIs, com destaque para o osimertinibe (Tagrisso®, AstraZeneca), um potente e seletivo inibidor para receptores que abrigam mutações de sensibilização e de resistência, principalmente a mutação T790M presente no éxon 20 do *EGFR* de pacientes com CPCNP cuja doença progrediu quando em uso de, ou após a terapia com EGFR-TKIs (53). Desde 2017, o osimertinibe era uma droga aprovada e disponível no Brasil indicada para o tratamento de segunda de pacientes com CPCNP localmente avançado ou metastático, positivo para mutação *EGFR* T790M, mutação secundária decorrente da primeira etapa da terapia com EGFR-TKIs. Em 2018, com a Resolução nº 927/2018, o Brasil foi o pioneiro no mundo, aprovando a ampliação de uso do osimertinibe uma vez ao dia, para o tratamento de primeira linha de pacientes com câncer metastático ou localmente avançado que, comprovadamente por teste laboratorial, que apresentam tumor com mutações de deleções do éxon 19 ou substituição do éxon 21 (L858R) do *EGFR* (54).

No entanto, destaca-se que, pacientes com CPCNP avançado com *EGFR* selvagem, o uso de EGFR-TKIs não é uma opção de tratamento de segunda linha após o insucesso de um esquema baseado em platina (51, 55).

2.7 A Transição Epitélio-Mesenquimal (EMT) e o câncer

A transição epitélio-mesenquimal (EMT) é um processo biológico em que a célula epitelial, por meio de várias alterações bioquímicas, perde progressivamente sua identidade e

morfologia celular assumindo um fenótipo de célula mesenquimal, que inclui aumento da capacidade migratória, de invasão e de resistência à apoptose (26).

As células epiteliais são células aderentes que formam camadas coerentes, as quais estão intimamente ligadas por complexos de adesão intercelular em suas membranas laterais. Elas também exibem polaridade ápico-basal, com uma membrana localizada basalmente a qual separa o epitélio de outros tecidos (57). As células epiteliais podem ser identificadas usando uma infinidade de marcadores de superfície celular, mais notavelmente a E-caderina, mas também citoqueratinas e claudinas (58).

Em contraste, as células mesenquimais são células não polarizadas e desprovidas de junções intercelulares, de modo que podem se mover como células individuais por toda a matriz extracelular, conforme ilustrado pela **Figura 7** (57). As células mesenquimais podem ser identificadas por marcadores de superfície celular de N-caderina, fibronectina e vimentina (58).

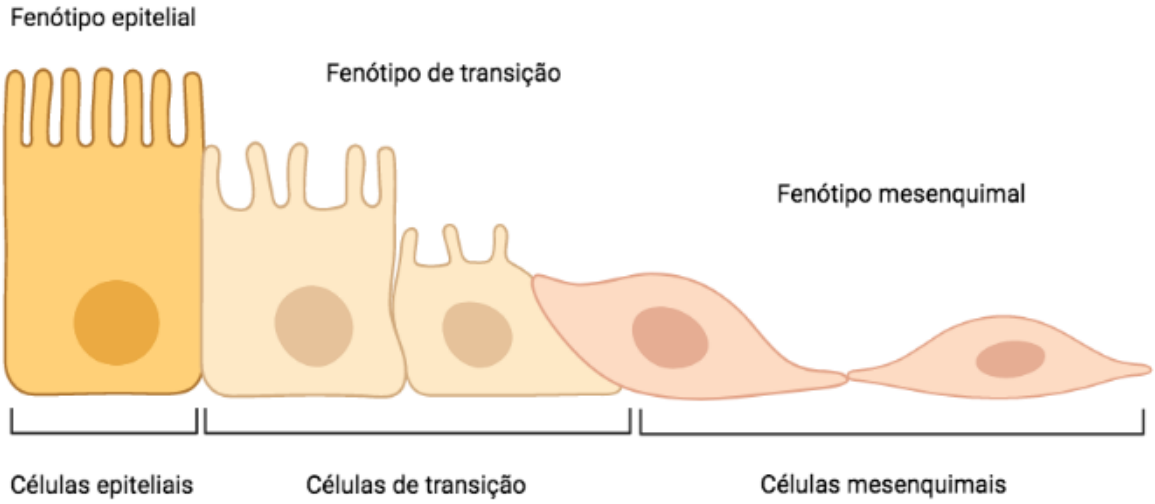


Figura 7: A EMT envolve uma transição funcional de células epiteliais polarizadas em células mesenquimais secretoras de componentes móveis.

Fonte: Próprio autor - criado em BioRender

A rota inversa é conhecida como transição mesênquimo-epitelial (MET). Tanto a EMT quanto a MET estão envolvidas em muitos processos normais de desenvolvimento celular como, por exemplo, no desenvolvimento embrionário, de cicatrização do tecido, e na

homeostase (58). Por exemplo, no desenvolvimento embrionário, as células da crista neural sofrem EMT, a fim de migrar para longe do tubo neural para se diferenciarem em osso, músculo liso, neurônios periféricos, glia, bem como melanócitos (59). Na fase adulta, este processo é reativado durante o processo de cicatrização de feridas, no qual células epiteliais diferenciam-se em miofibroblastos e reconstroem a matriz extracelular a fim de facilitar a contração da ferida (60). No entanto, a EMT também pode ser ativada em processos patológicos, como por exemplo, na cicatrização de feridas, no qual o excesso de tecido conjuntivo fibroso pode levar à fibrose de órgãos. A EMT também pode ser ativada por um conjunto de células cancerosas, para permitir que invadam e formem metástases em locais distantes (59, 61).

Desta forma, a EMT pode ser classificada em três subtipos com base nas condições biológicas em que ocorrem e nas consequências associadas, sendo elas: EMT tipo1, que ocorre durante a embriogênese e no desenvolvimento dos órgãos; EMT tipo 2, que ocorre durante a reparação tecidual e fibrose; e a EMT tipo 3, que ocorre na progressão tumoral e metástase, de acordo com a ilustração da **Figura 8**.

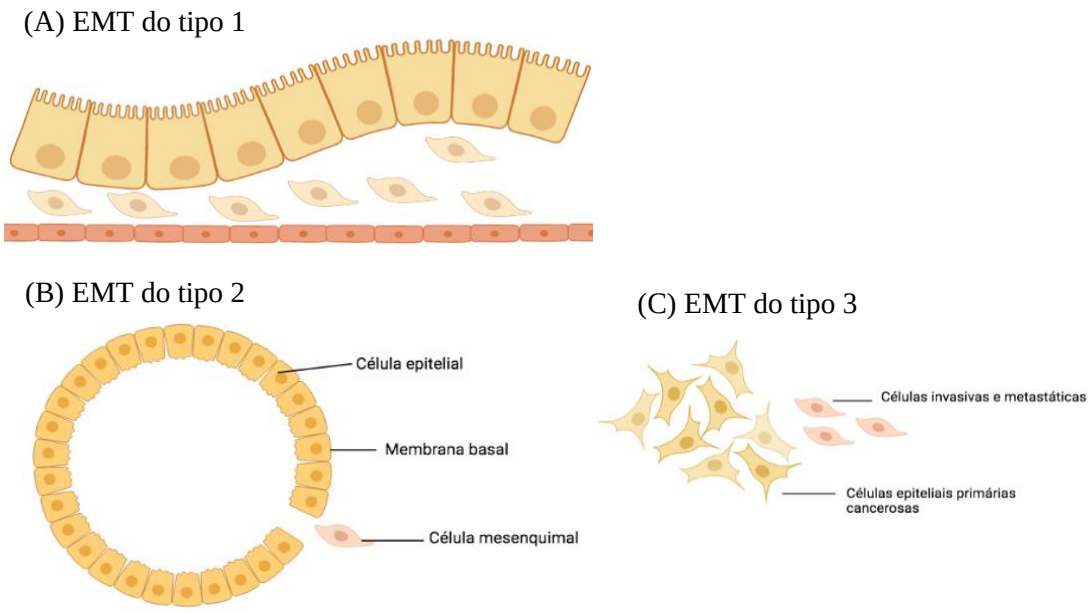


Figura 8: Esquema ilustrativo dos diferentes tipos de EMT. (A) EMT do tipo I, relacionada ao desenvolvimento embrionário; (B) EMT do tipo II relacionada ao processo de cicatrização e fibrose, e (C) EMT do tipo III associada à progressão tumoral e metástases.

Fonte: Próprio autor - criado em BioRender.

A metástase é responsável por cerca de 90% das mortes associadas ao câncer, reforçando a importância da intervenção no processo da EMT. A EMT é regulada por múltiplos fatores em diferentes níveis, incluindo sinalização celular, fatores transcricionais e pós-transcricionais.

Na sinalização celular, destacam-se várias moléculas de sinalização como as citocinas, tais como: membros da família de fatores de transformação do crescimento beta (TGF-β), família de fatores de crescimento de fibroblastos (FGF), fatores de crescimento epidérmico (EGF) e fatores de crescimento de hepatócitos (HGF), proteína morfogenética óssea (BMP), Sonic Hedgehog (SHH), a via de sinalização Notch, entre outros (58).

Dentre os fatores de transcrição (FT) relacionados ao processo EMT destacam-se *TWIST1*, *TWIST2*, *SNAIL1*, *SNAIL2*, *ZEB1* e *ZEB2*, os quais, como consequência de sua expressão, inibem diretamente biomarcadores da superfície das células epiteliais, principalmente, a expressão da E-caderina promovendo a transição para um estado mesenquimal. Já entre os fatores pós-transcricionais, destacam-se os microRNAs (miRNAs) e,

mais recentemente, os RNAs longos não codificantes. Os miRNAs são pequenos RNAs não codificantes que podem pós-transcricionalmente ajustar a expressão gênica ao direcionar a 3′UTR de mRNA, levando à sua desestabilização e degradação (62). Até o momento, mais de 130 miRNAs estão relacionados na regulação da EMT com destaque para os membros da família miR-200, que incluem: miR-200a, miR-200b, miR-200c, miR-141 e miR-42, os quais podem inibir a EMT ao direcionar as moléculas de mRNA dos fatores promotores de EMT, resultando em sua degradação do transcrito ou repressão translacional (63).

Já os RNAs longos não codificantes (lncRNAs), isto é, RNAs não codificantes com mais de 200 nucleotídeos de comprimento, também estão envolvidos em uma infinidade de processos biológicos, incluindo o processo EMT. Entre suas funções no controle de EMT, eles podem regular as vias de sinalização, incluindo a de TGFβ, podem funcionar como RNAs endógenos competidores (ceRNAs) por miRNAs ou até mesmo influenciar a expressão de genes associados a EMT, incluindo os FTs de EMT (58).

3. JUSTIFICATIVA

Apesar dos avanços terapêuticos e de diagnóstico, o câncer de pulmão é uma doença molecularmente heterogênea, sendo fundamental a caracterização molecular para melhorar a compreensão da patogênese do tumor (64). A presente tese explorou alvos moleculares envolvidos na patogênese do tumor. Nos artigos que serão apresentados a seguir realizamos análises genéticas do tumor por meio da técnica de sequenciamento de nova geração (NGS).

O artigo 1 é intitulado "*Variants in Epithelial-Mesenchymal Transition and Immune Checkpoint Genes Are Associated With Immune Cell Profiles and Predict Survival in Non–Small Cell Lung Cancer".* A hipótese deste estudo era que variantes genéticas relacionadas à EMT e fatores imunológicos do hospedeiro podem predizer os resultados clínicos de pacientes com NSCLC.

O artigo 2 é intitulado "*Relevance of PD-L1 Non-Coding Polymorphisms on the Prognosis of a Genetically Admixed NSCLC Cohort*". Nossa hipótese era que variantes genéticas não codificantes de PD-L1 podem modular a função deste gene no NSCLC.

Por último, o artigo 3, intitulado "*Clinical outcome of Brazilian patients with non-small cell lung cancer in early stage harboring rare epidermal growth factor receptor mutations*" no qual investigamos a frequência de mutação no gene *EGFR*, com o objetivo de explorar o resultado clínico de pacientes com CPNPC ressecado em estágio inicial portadores de mutações de significado clínico incerto.

4. OBJETIVOS

GERAL

Avaliar o perfil de alteração genética em carcinomas de pulmão de células não pequenas utilizando a técnica de Sequenciamento de Nova Geração (NGS).

ESPECÍFICOS

Traçar o perfil imunológico do CPCNP utilizando a imunofluorescência multiplex e correlacionar os achados com o perfil de alteração genética relacionado à EMT;

Investigar a influência de variantes não-codificantes do gene *PD-L1* em amostras de CPCNP e correlacionar com expressão proteica e as características clinicopatológicas;

Estimar a frequência de mutação no gene *EGFR*, com destaque para as mutações de significado clínico incerto no CPCNP.

5. RESULTADOS

5.1 Artigo 1: citação completa (ANEXO 1)

Edwin Roger Parra, Mei Jiang, Juliana Machado-Rugolo, Lygia Bertalha Yaegashi, Tabatha Prieto, Cecília Farhat, Vanessa Karen de Sá, Maria Aparecida Nagai, Vladmir Cláudio Cordeiro de Lima, Tereza Takagaki, Ricardo Terra, Alexandre Todorovic Fabro, Vera Luiza Capelozzi; Variants in Epithelial-Mesenchymal Transition and Immune Checkpoint Genes Are Associated With Immune Cell Profiles and Predict Survival in Non–Small Cell Lung Cancer. *Arch Pathol Lab Med* 1 October 2020; 144 (10): 1234–1244.

doi: https://doi.org/10.5858/arpa.2019-0419-OA

Fator de impacto = 5.686

Variants in Epithelial-Mesenchymal Transition and Immune Checkpoint Genes Are Associated With Immune Cell Profiles and Predict Survival in Non–Small Cell Lung Cancer

Edwin Roger Parra[1], Mei Jiang[1], Juliana Machado-Rugolo[2], Lygia Bertalha Yaegashi[2], Tabatha Prieto[2], Cecília Farhat[2], Vanessa Karen de Sá[3], Maria Aparecida Nagai[4], Vladmir Cláudio Cordeiro de Lima[3], Tereza Takagaki[5], Ricardo Terra[6,7], Alexandre Todorovic Fabro[8], Vera Luiza Capelozzi[2]

[1] Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, EUA.

[2] Department of Pathology and Laboratory of Genomics and Histomorphometry, Faculty of Medicine, University of São Paulo, São Paulo, Brazil

[3] Medical Oncology Department and Translational Immune-Oncology Group, A. C. Camargo Cancer Center, São Paulo, Brazil

[4] Department of Oncology, Clinicas Hospital, Faculty of Medicine, São Paulo State University, São Paulo, Brazil

[5] Division of Pneumology, Heart Institute (Incor), São Paulo, Brazil

[6] Department of Thoracic Surgery, Institute of Cancer of São Paulo, São Paulo, Brazil

[7] Department of Thoracic Surgery, Heart Institute (Incor), São Paulo, Brazil

[8] Department of Pathology and Legal Medicine, Ribeirão Preto School of Medicine, University of São Paulo, Ribeirão Preto, Brazil

Abstract

Context: Identification of gene mutations that are indicative of epithelial-mesenchymal transition and a noninflammatory immune phenotype may be important for predicting response to immune checkpoint inhibitors.
Objective: To evaluate the utility of multiplex immunofluorescence for immune profiling and to determine the relationships among tumor immune checkpoint and epithelial-mesenchymal transition genomic profiles and the clinical outcomes of patients with nonmetastatic non– small cell lung cancer.
Design: Tissue microarrays containing 164 primary tumor specimens from patients with stages I to IIIA non– small cell lung carcinoma were examined by multiplex immunofluorescence and image analysis to determine the expression of programmed death ligand-1 (PD-L1) on malignant cells, CD68+ macrophages, and cells expressing the immune markers CD3, CD8, CD57, CD45RO, FOXP3, PD-1, and CD20. Immune phenotype data were tested for correlations with clinicopathologic characteristics, somatic and germline genetic variants, and outcome.
Results: A high percentage of PD-L1^{+} malignant cells was associated with clinicopathologic characteristics, and high density of CD3^{+}PD-1^{+} T cells was associated with metastasis, suggesting that these phenotypes may be clinically useful to identify patients who will likely benefit from immunotherapy. We also found that *ZEB2* mutations were a proxy for immunologic ignorance and immune tolerance microenvironments and may predict response to checkpoint inhibitors. A multivariate Cox regression model predicted a lower risk of death for patients with a high density of CD3^{+}CD45RO^{+} memory T cells, carriers of allele G of CTLA4 variant rs231775, and those whose tumors do not have *ZEB2* mutations.
Conclusions: Genetic variants in epithelial-mesenchymal transition and immune checkpoint genes are associated with immune cell profiles and may predict patient outcomes and response to immune checkpoint blockade

Keywords: Immunoprofiling; lung cancer; multiplex immunofluorescence; image analysis; next-generation sequencing; histologic types; genetic analysis; genetic variants

Introduction

Non-small cell lung cancer (NSCLC) accounts for about 85% of all lung cancers. The prognosis of patients with NSCLC is highly dependent on the tumor-node-metastasis (TNM) stage; as the volume of disease increases, survival declines dramatically[1]. Surgery is the standard treatment for patients with stage I and stage II NSCLC and for selected patients with stage IIIA disease with the possibility of complete tumor resection. The addition of adjuvant cisplatin-based chemotherapy to surgery improved 5-year survival rates by 5% to 10%, but no significant therapeutic innovation has been established thereafter, and the overall 5-year survival rate remains below 50%.

Over the past few years, immunotherapy with immune checkpoint inhibition has led to remarkable progress in non-oncogene addicted metastatic NSCLC[2]. Most of these new immunotherapeutic approaches for advanced NSCLC are now being studied in the context of nonmetastatic NSCLC[3]. However, not all NSCLC patients receive the same benefit from immunotherapy as a first or second line of treatment. To date, programmed death ligand-1 (PD-L1) is the only validated biomarker for immunotherapy response in NSCLC patients. However, this biomarker is not sufficiently robust and poses many challenges. For example, some patients with more than 50% PD-L1þ tumor cells do not have a response to anti–programmed death receptor-1 (PD-1)/PD-L1 treatment; conversely, some patients whose tumors do not express PD-L1 see good responses.[4-6]

Some studies have suggested that the heterogeneity of PD-1/PD-L1 axis expression and response to PD-1/PD-L1 treatment in NSCLC depends on alterations in other factors, such as immune evasion mechanisms and the tumor immune microenvironment. Expression of genes involved in the epithelial-mesenchymal transition (EMT) and immune checkpoints are also known to control cancer progression[7-16]. However, the associations among these factors and their association with immunotherapy response remain to be determined.

We hypothesized that genetic variants that are indicative of EMT and host immune factors predict patients' outcomes and responses to immune checkpoint blockade. Identification of mutations that can serve as a proxy for a noninflammatory immune phenotype would further enable us to predict response to immune checkpoint blockade. Such information would be particularly important in clinical settings where next-generation sequencing (NGS) is available, but comprehensive studies of tumor-associated immune cells (TAICs) are not. Bearing this in mind, we determined the extent to which immune and genomic profiles predicted clinical outcomes of patients with NSCLC.

PATIENTS AND METHODS

Cases and Specimens

We obtained archival formalin-fixed, paraffin-embedded histologic tumor sections from 164 patients diagnosed with NSCLC who had undergone surgical resection between January 1, 1995, and December 31, 2015, and who had not received neoadjuvant therapy. Patients had been treated at the Hospital das Clinicas of the State University of São Paulo Medical School, the Heart Institute of the University of São Paulo, the Cancer Institute of São Paulo, and the A.C. Camargo Cancer Center, São Paulo, Brazil. The histologic diagnoses were reviewed by 2 experienced lung pathologists, who judged their accuracy based on the World Health Organization 2015 classification system[17]. The specimens included 94 adenocarcinomas, 51 squamous cell carcinomas (SCCs), and 19 large cell carcinomas (LCCs). Tissue microarrays were constructed from the primary resected tumors: 3 cores (diameter, 1.5 mm) of representative tumor tissue were taken from a representative formalin-fixed, paraffin-embedded block from each patient case. Clinicopathologic data, including pathologic TNM staging, which was determined according to the guidelines of the International Association for the Study of Lung Cancer, 8th edition,[1] were obtained from medical records (**Table 1)**.

Our study received approval from the institutional review boards of the institutions involved in the project (process number H-1404-100-572). A waiver of the requirement for informed consent was obtained from the review boards of all participating institutions.

Multiplex Immunofluorescence Staining

Multiplex immunofluorescence (mIF) staining was performed using methods similar to previously described and validated ones[18]. Consecutive 4-lm-thick tissue microarray sections were stained using an automated staining system (BOND-RX, Leica Biosystems, Buffalo Grove, Illinois) using 2 different panels. **Panel 1** consisted of pancytokeratin AE1/ AE3 (dilution 1:300; epithelial cell positive, Dako, Carpinteria, California), PD-L1 (dilution 1:100; clone E1L3N, Cell Signaling Technology, Danvers, Massachusetts), PD-1 (dilution 1:250; clone EPR4877-2, Abcam, Cambridge, Massachusetts), CD3 (dilution 1:100; T-cell lymphocytes, Dako), CD8 (dilution 1:20; cytotoxic T cells, clone C8/144B, Thermo Fisher Scientific, Waltham, Massachusetts), and CD68 (dilution 1:450; macrophages, clone PG-M1, Dako). **Panel 2** consisted of AE1/AE3 (epithelial cell-positive, dilution 1:300; Dako), CD3 (dilution 1:100; T cell lymphocytes, Dako), CD57 (dilution 1:40; natural killer cells, clone HNK-1, BD Biosciences, San Jose, California), CD45RO (ready to use; memory T cells, clone UCHL1, Leica Biosystems), FOXP3 (dilution 1:50; regulatory T cells, clone 206D, BioLegend, San Diego, California), and CD20 (dilution 1:100; B-cell lymphocytes, Dako). All the markers were stained in sequence using their respective fluorophores, which were contained in the Opal 7 kit (Cat #NEL797001KT, Akoya Biosciences/PerkinElmer, Waltham, Massachusetts) after IF validation to obtain a uniform, specific, and correct signal across all channels and ensure a well-balanced staining pattern during the multiplex staining[18]. The correct signal from each fluorophore was also defined and optimized between 10 and 30 counts of intensity to maintain a good balance and similar thresholds of intensity across all the antibodies. In parallel, to detect possible variations in staining and optimal separation of the signal, positive and negative

(autofluorescence) controls were included during the staining. Autofluorescence controls with an expected spectrum of 488 nm can accurately remove the autofluorescence from all the labeling signals during the analysis process. The stained slides were scanned using a multispectral microscope (Vectra Polaris 3.0 imaging system, Akoya Biosciences/PerkinElmer) under fluorescence conditions (**Supplementary Figure 1**).

mIF Quantitation

Multispectral images of tumor sections from each core were analyzed with inForm 2.2.1 software (Akoya Biosciences/PerkinElmer). Individual cells, which were defined by nuclear staining and identified by the inForm cell segmentation tool, were used in a phenotype pattern-recognition learning algorithm to characterize the colocalization of the cell populations labeled with 2 panels.[19] Panel 1 labeling was as follows: (1) malignant cells (MCs) with the AE1/AE3+ marker, including those with and without PD-L1 expression (AE1/AE3$^+$PD-L1$^+$ and AE1/AE3$^+$PD-L1, respectively); (2) T lymphocytes, including CD3$^+$ and pan–T-cell markers (cytotoxic T cells [CD3$^+$CD8$^+$], antigen-experienced T cells [CD3$^+$PD-1$^+$], and other CD3$^+$ T cells); (3) CD68$^+$ macrophages; and (4) macrophages expressing PD-L1 (CD68$^+$PDL1$^+$). Panel 2 labeling was as follows: (1) MCs with the AE1/ AE3$^+$ marker; (2) T lymphocytes, including CD3þ and pan– T-cell markers (memory T cells [CD3$^+$CD45RO$^+$], natural killer T cells [CD3$^+$CD57$^+$], memory/regulatory T cells [CD3$^+$CD45RO$^+$FOXP3$^+$], and other CD3$^+$ T cells); and (3) B-cell lymphocytes (CD20$^+$).

The individual cell phenotype report produced by the inForm software was processed using Microsoft Excel 2010, and a final summary of the data, which contained the median number of cells/mm2 for each phenotype and the median percentage of macrophages and MCs expressing PD-L1, was created for statistical analysis. If the percentage of MCs or macrophages expressing PD-L1 was greater than the median value, PD-L1 expression was considered positive. If the percentage of macrophages or MCs expressing PD-L1 in a sample was lower

than or equal to the median, PD-L1 expression was considered negative. Likewise, if the density of TAICs in a cell population was greater than the median, it was considered high density; if the density was less than or equal to the median, it was considered low.

In addition, as proposed by Teng and colleagues,[20] we identified 4 different types of tumor microenvironments on the basis of the density of CD3+ tumor-infiltrating lymphocytes (TILs) (a density greater than the median was considered high) and the PD-L1 expression level of MCs. The 4 microenvironment types were type I, adaptive immune resistance (PD-L1+/TIL+); type II, immunologic ignorance ($PD\text{-}L1^{-}/TIL^{-}$); type III, intrinsic induction ($PDL1^{+}/TIL^{-}$); and type IV, tolerance ($PD\text{-}L1^{-}/TIL^{+}$).

Genomic Analysis

Following previously published procedures,[21] we extracted DNA from fresh-frozen lung cancer tissue obtained from 70 of the 164 patients in our study; 33 of the fresh-frozen samples were from patients with adenocarcinoma, 24 were from patients with SCC, and 13 were from patients with LCC. To study the tumor genomic landscape from different angles, we used targeted NGS to evaluate single-nucleotide variants and short deletions/insertions (indels) in NSCLC samples, including 13 genes, 121 targets, and 372 amplicons associated with these genes. We verified the results with DNA sequencing analysis with an Illumina TruSeq Custom Amplicon Panel (**Supplemental Table 1**). The genes examined were grouped according to their functions as immune checkpoints (*CD274* [*PD-L1*], *CD276*, C*TLA4*, *PDCD1LG2* [*PD-L2*], *LAG3*, and *VTCN1* [*B7H4*]) or EMT genes (*CD44*, *CDH1* [E-cadherin], *TGFβ1*, *VIM*, *MMP2*, *ZEB1*, and *ZEB2*). Libraries were sequenced with an Illumina MiSeq platform using MiSeq v2 (300 cycles). Primary data analysis was performed using the TruSeq Amplicon pipeline (alignment to reference genome hg19) via the Basespace Sequence Hub. Illumina VariantStudio v2.2 software was used for filtering and variant annotation. Variants with frequency lower than 1% in the study population were considered single-nucleotide variants,

and variants with frequency higher than 1% were defined as somatic variants. Filtered retained variants had to have a total coverage depth of 100 reads or more and a variant allele frequency of at least 5% to be considered of adequate quality to be interpreted. The potential pathogenicity of the allele variants was evaluated with 3 publicly available algorithms (Polyphen2, Sift, ClinVar). The distribution of mutations in EMT and immune checkpoint genes in our study cohort was compared with data from The Cancer Genome Atlas. All mutations were tested for correlations with patients' clinicopathologic characteristics, TAIC levels, and outcomes.

Statistical Analysis

The $\chi 2$ test or Fisher exact test was used to examine differences in categorical variables, and the Wilcoxon rank sum test and Kruskal-Wallis test were used to detect differences in continuous variables among groups of patients. In addition, a general linear regression model was used to test the relationships between continuous variables and several other variables, and the residuals were examined to ensure that they were approximately normally distributed. The overall survival (OS) distributions for the patients were estimated using the Kaplan-Meier method. Overall survival was defined as the interval from surgery to death and was estimated using the Kaplan-Meier method, and the log-rank test was used to compare survival among the groups of interest. Regression analysis of the OS data was performed using the Cox proportional hazards model. Variables shown by univariate analysis to be significantly associated with survival were entered into a Cox proportional hazards regression model for multivariate analysis. The statistical software programs SPSS (version 22, IBM, Armonk, New York) and S-Plus (version 8.04, TIBCO, Palo Alto, California) were used to perform the analyses. P values less than or equal to .05 were deemed statistically significant.

RESULTS

Clinicopathologic Characteristics

The clinical characteristics of the 164 patients in our cohort are summarized in **Table 1** by histologic tumor type. Similar distributions of age and sex were found among the histologic types. Of the 106 patients with a history of tobacco smoking, 51 (48%) had adenocarcinoma, 48 (45%) had SCC, and 7 (7%) had LCC. The median follow-up duration was 43 months (range, 6–180 months). During the follow-up period, disease progression (local recurrence and distant metastasis to liver, bones, lungs, or brain) occurred in 30 of the 164 patients: 16 (53%) with adenocarcinoma, 12 (40%) with SCC, and 2 (7%) with LCC. All patients with distant metastases also had lymph node metastases. Of the 164 patients, 34 (21%) received adjuvant therapy, and 56 (34%) had died of their disease by last follow-up.

Table 1. Clinical characteristics of 164 patients with non-small cell lung cancer.

Characteristic	ADC (n=94)	SCC (n=51)	LCC (n=19)
Age, median, years	66	67	59
Sex, n (%)			
Female	46 (49)	19 (37)	6 (32)
Male	48 (51)	32 (63)	13 (68)
Tobacco history, n (%)			
No	19 (20)	1 (2)	1 (5)
Yes	51 (54)	48 (94)	7 (37)
Unknown	24 (26)	2 (4)	11 (58)
Tumor size, median, cm	3.0	3.2	5.0
Tumor status†, n (%)			
T_1	29 (30)	14 (27)	6 (32)
T_2	47 (51)	30 (59)	13 (68)
T_3	15 (16)	5 (10)	0 (0)
T_4	3 (3)	2 (4)	0 (0)
Nodal status†, n (%)			
N_0	82 (87)	41 (80)	10 (53)
N_1	12 (13)	10 (20)	9 (47)
IASLC pathologic stage†, n (%)			
I	44 (47)	30 (59)	6 (31.6)
II	29 (31)	11 (21)	5 (26)

IIIA	21 (22)	10 (20)	4 (21)
Disease recurrence, n (%)	16 (17)	12 (24)	2 (11)
Adjuvant treatment, n (%)			
No	74 (79)	41 (80)	15 (79)
Yes	20 (21)	10 (20)	4 (21)
Patients censored for survival analysis at last time of follow-up, n (%)	32 (34)	15 (29)	9 (47)

Abbreviations: IASLC, International Association for the Study of Lung Cancer; LCC, large cell carcinoma; SCC, squamous cell carcinoma. † IASLC staging guidelines, 8th edition.

PD-L1 ***Expression in MCs and Macrophages***

The median cell densities of MCs and macrophages by histologic tumor type are shown in **Table 2**. Image analysis and mIF revealed no significant differences in the median densities of PD-L1$^+$ MCs among the histologic types: the median density was 0.39 cells/mm^2 for adenocarcinoma, 0.48 cells/mm^2 for SCC, and 1.59 cells/mm^2 for LCC (P =.60) (**Table 2**). Using the median percentage of MCs expressing PD-L1 across all the samples as a cutoff, we identified 45 of 94 (47%) PD-L1$^+$ adenocarcinoma samples, 25 of 51 (49%) PD-L1$^+$ SCC samples, and 6 of 19 (31%) PDL1$^+$ LCC samples (Figure, A through F). For all 3 histologic types, the density of CD68$^+$PD-L1$^+$ TAMs was higher than that of PD-L1$^+$ MCs.

Table 2. Median densities of various immune marker–expressing cells according to histologic tumor type (n=164).

Marker	ADC (n=94)	SCC (n=51)	LCC (n=19)	*P**
	Median Cell Density (cells/mm^2)			
Panel 1				
MCs (AE1/AE3+)	4780.25	4311.02	4366.56	.55
MCs (AE1/AE3+) PD-L1+	0.39	0.48	1.59	.60
Total CD3+	432.90	246.54	423.12	.02
CD3+CD8+	112.71	68.56	107.55	.04
CD3+PD-1+	59.21	23.23	79.50	.01
CD68+	82.09	69.76	100.60	.36
CD68+PD-L1+	2.03	1.11	3.50	.85
Panel 2				
MCs (AE1/AE3+)	5294.40	5407.66	6570.75	.57
Total CD3+	881.43	687.63	694.81	.55
CD3+CD45RO+	534.29	342.31	592.97	.06
CD3+CD57+	20.33	30.01	21.69	.60
CD3+CD45RO+FOXP3+	145.80	83.14	307.63	.36
CD20+	45.69	41.14	24.84	.41

Abbreviations: ADC, adenocarcinoma; LCC, large cell carcinoma; MCs, malignant cells; PD-L1, Programmed death-ligand 1, SCC, squamous cell carcinoma.

* The Kruskal-Wallis test was used to detect differences in continuous variables between groups of patients. A *P* value ≤ .05 was considered statistically significant.

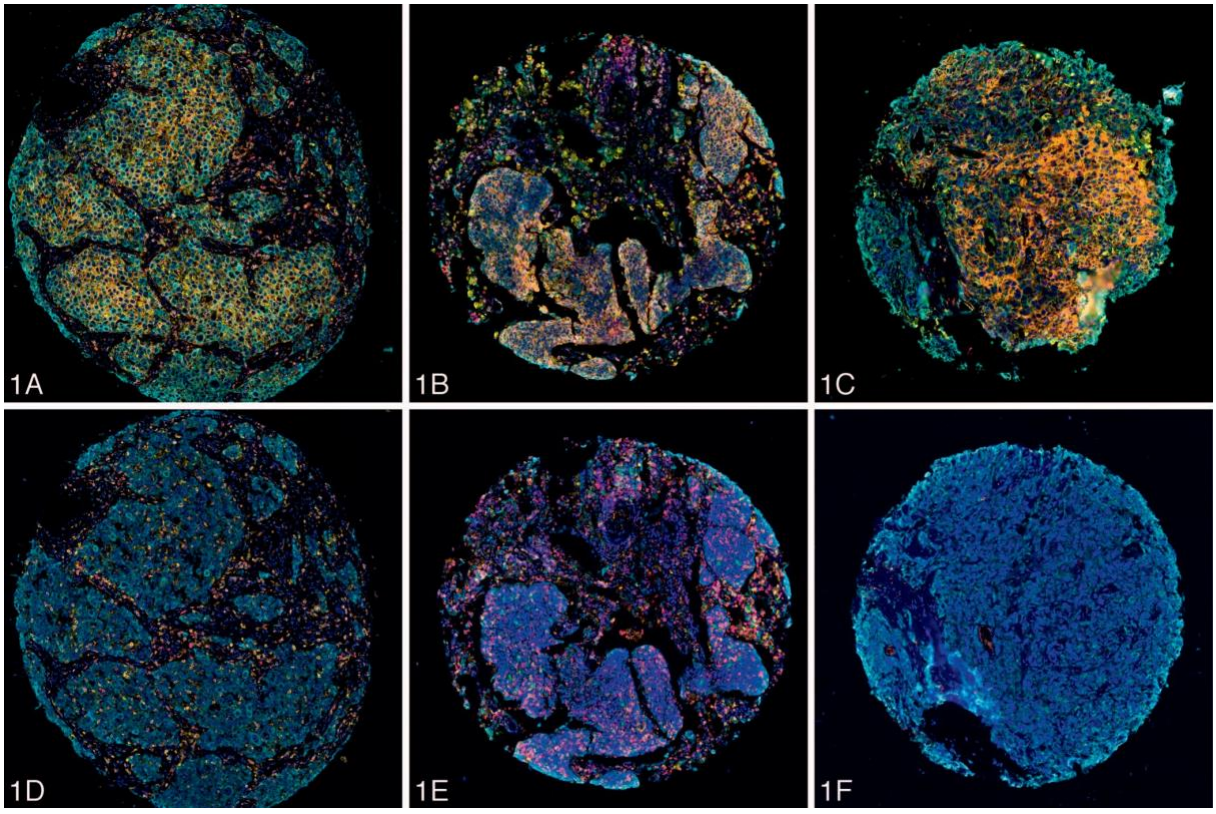


Figure 1: Multiplex immunofluorescence images of representative non–small cell lung cancer core tumor sections analyzed for panel 1 (pancytokeratin, cyan; PD-L1, orange; CD3, red; CD8, pink; PD-1, green; CD68, yellow; DAPI, blue) and panel 2 (pancytokeratin, cyan; CD3, red; CD8, pink; CD45RO, yellow; GranzymeB, orange; FOXP3, green; DAPI, blue) markers. The images reflect the variations in cell phenotypes observed in adenocarcinoma (A and D, panels 1 and 2, respectively), squamous cell carcinoma (B and E, panels 1 and 2, respectively), and large cell carcinoma (C and F, panels 1 and 2, respectively), especially in regard to PD-L1 expression by malignant cells.

Clinicopathologic correlations with PD-L1 and TAICs

Certain immune cell phenotypes were significantly associated with patients' clinicopathologic characteristics, as shown in **Supplemental Table 2**. A general linear regression model demonstrated that low PD-L1 expression on MCs was significantly associated with tobacco use (P = .04) and adenocarcinoma (P = .04). Low density of natural killer T cells (CD3+CD57+) was also associated with tobacco use (P = .04). Low density of cytotoxic T cells (CD3+CD8+) was associated with LCC (P = .04). In addition, low density of memory T cells (CD3+CD45RO+) was associated with female sex (P = .01) and age 65 years or older (P = .04).

Importantly, high density of antigen-experienced T cells (CD3+PD-1+) was associated with brain metastasis (P = .04), and, curiously, high density of B cells (CD20+) was significantly more common among patients who did not receive adjuvant therapy than among those who did (P= .04). All these associations remained statistically significant after allowing for the contribution of other clinicopathologic characteristics, as determined by multivariate analysis.

Characterization of Tumor Immune Microenvironments

We observed significantly higher densities of CD3+ T cells (P = .02), CD3+CD8+ cytotoxic T cells (P = .04), and CD3+PD-1+ antigen–experienced T cells (P = .01) in adenocarcinoma and LCC than in SCC (**Table 2**). No other statistically significant differences were found among the histologic types.

By examining the PD-L1 expression of MCs in conjunction with the density of CD+þ TILs, as proposed by Teng and colleagues,[20] among the different histologic types, we identified all 4 types of tumor microenvironments in our NSCLC samples (**Table 3**). Type I, adaptive immune resistance (PD-L1+/TIL+), was observed in 11 of 94 (12%) adenocarcinoma samples, 7 of 51 (14%) SCC samples, and 4 of 19 (21%) LCC samples. Type II, immunologic ignorance (PD-L1/TIL), was found in 39 of 94 (41%) adenocarcinoma samples, 28 of 51 (55%) SCC samples, and 7 0f 19 (37%) LCC samples. Type III, intrinsic induction (PD-L1+/TIL), was found in 3 of 94 (3%) adenocarcinoma samples, 4 of 51 (88%) SCC samples, and 3 of 19 (16%) LCC samples. Type IV, tolerance (PD-L1/TIL+), was found in 41 of 94 (42%) adenocarcinoma samples, 12 of 51 (24%) SCC samples, and 5 of 19 (26%) LCC samples.

Table 3. Distribution of types of immune microenvironments in 164 non-small cell lung cancer specimens[a]

		Adenocarcinoma, No (%) (n=94)	
		TILs	
		Negative	Positive
PD-L1[b]	Negative	39 (41)	41 (43)
	Positive	3 (3)	11 (12)
		Squamous cell carcinoma, No (%) (n=51)	
		TILs	
		Negative	Positive
PD-L1[b]	Negative	28 (55)	12 (24)
	Positive	4 (8)	7 (14)
		Large cell Carcinoma, No (%) (n=19)	
		TILs	
		Negative	Positive
PD-L1[b]	Negative	7 (37)	5 (26)
	Positive	3 (16)	4 (21)

Abbreviations: PD-L1, programmed death ligand-1; TIL, tumor-infil- trating lymphocyte.
[a] The tumor immune microenvironments are based on the percentage of malignant cells expressing PD-L1 and the density of CD3+ TILs.
[b] PD-L1 expression was considered negative if the percentage of malignant cells expressing PD-L1 on the cell membrane was less than or equal to the median percentage for that histologic type and positive if greater than the median.
[c] TILs were considered negative if the density of CD3+ cells was less than or equal to the median density for that histologic type and positive if greater than the median.

Correlations between genomic profile, clinicopathologic characteristics, and immunophenotypes

Using NGS, we examined the occurrence of variants in several cancer-related genes in fresh-frozen lung cancer tissue specimens obtained from 70 of the 164 NSCLC patients included in our study, and we correlated the mutation status of these genes with patients' clinicopathologic characteristics and immune cell populations. The distribution of mutations in EMT and immune checkpoint genes in our study cohort was compared with data from The Cancer Genome Atlas (**Supplemental Table 3**). We found similar frequencies of mutations in immune checkpoint and EMT genes in our NSCLC cohort and the Cancer Genome Atlas cohort. The most frequently mutated genes in our cohort were *ZEB2* (n=10; 14%), MMP2 (n=4;

6%), *ZEB1* (n = 2; 3%), *CDH1* (n = 2; 3%), and *CD44* (n = 2; 3%), all of them involved in EMT. On multivariate analysis, *CD44* mutations were significantly associated with tobacco use (P < .05). The small number of mutations in *TGFβ1* and *VIM* did not allow for multivariate analysis to test for associations with clinicopathologic characteristics. The genotype and allele frequencies of 5 variants, rs7854303 (*PDCD1LG2* [*PD-L2*]), rs2297136 (*CD274* [*PDL1*]), rs870849 (*LAG3*), rs10754339 (*VTCN1* [*B7H4*]), and rs231775 (*CTLA4*), in our cohort are summarized in Table 4. Only genotype CC and the C allele of *PD-L2* variant rs7854303 were identified in all patients. On multivariate analysis, the only significant difference in clinicopathologic characteristics by genotype was that the AA genotype of the *CTLA4* rs231775 variant was more common in stage II and III than in stage I tumors (P = .02) (**Supplemental Table 2**). No association was found between rs2297136 (*PD-L1*), rs870849 (*LAG3*), or rs10754339 (*VTCN1*) and any clinicopathologic characteristics. Our investigation of the association between genomic profiles and immune cell profiles revealed significant correlations between high density of CD3+PD-1+ antigen– experienced T cells and the G allele (GG+GA) of *CTLA4* rs231775 (R = 0.33; P = .01), high density of CD3+ TILs and *ZEB1* mutation (R =0.27; P= .04), and high density of CD3+CD8+ cytotoxic T cells and *ZEB1* mutation (R =0.34; P = .01) (**Supplemental Table 4**).

Survival analysis

Preliminary examination of Kaplan-Meier survival curves (data not shown) demonstrated that in this study, patients with pathological stages II and III had approximately the same hazard for survival with a median survival time equal to 40 months for both groups. Thus, we coded overall pathological stage as a single dummy variable with a value of 0 for stage I and a value of 1 for stages II and III. The results of the Cox proportional hazards regression models appear in **Tables 4** and **5**.

The following clinicopathologic variables were significantly associated with poor OS: female sex, tobacco use, tumor size greater than 4.5 cm, N_1 status, tumor stage higher than II, SCC histotype type, and absence of adjuvant treatment (**Table 5**). We also found that lower-than-median densities of CD3+CD8+, CD3+CD45RO+, and CD3+CD45RO+FOXP3+ T cells in the tumor microenvironment predicted shorter OS, as did *ZEB2*-mutant tumors. In addition, *CTLA4* variant rs231775 with AA genotype was associated with shorter OS, whereas allele G was significantly associated with longer OS (adjusted hazard ratio=2.091, 95% CI=0.233–2.219, P=.05) (**Table 4**). No other variables included in the multivariate analysis were significantly associated with survival. Multivariate analysis controlling for CD3+CD8+ T cell density, CD3+CD45RO+ T cell density, *CTLA4* variants, and *ZEB2* mutations revealed that tumor size, N stage, histologic type, and adjuvant treatment significantly predicted survival time. However, when CD3+CD45RO+ T cells, *CTLA4* rs231775 G allele carrier genotype, and *ZEB2* wild type were included as covariates, their relationship to survival was much stronger. Whereas the overall likelihood ratio of the Cox model using tumor size, N staging, histologic type, and treatment was just 21.83, the likelihood ratio with stage, CD3+CD45RO+ T cells, *CTLA4* variant rs231775 G allele carrier genotype, and *ZEB2* wild-type tumors was 31.18.

Table 4. Genotype distributions of genetic variants in patients with NSCLC and association with overall survival (OS).

Genes	Chromosome Position	Genetic Variant	Genotype	No. of case (%)	Univariate Analysis[a]			Multivariate Analysis[b]	
					HR[c] (95% CI)	HR[c]	*P* v	HR[c] (95% CI)	*P*
PDCD1LG2 (PD-L2)	9p24.1	rs7854303	TT	0 (0)	1.0 (reference)				
			TC	0 (0)					
			CC	70 (100)					
			C allele carrier	70 (100)	----				
CD274 (PD-L1)	9p24.1	rs2297136	GG	15 (21.4)	1.0 (reference)				
			GA	25 (35.7)					
			AA	30 (42.9)					
			A allele carrier	55 (78.6)	0.57 (0.20-1.63)	-.056	.29		
LAG3	12p13.31	rs870849	TT	9 (12.9)	1.0 (reference)				
			TC	32 (45.7)					
			CC	29 (41.3)					
			C allele carrier	61(87.1)	2.65 (0.73-9.57)	0.97	.13	0.18 (0.00-0.98)	.42
VTCN1 (B7H4)	1p13.1-p12	rs10754339	GG	4 (5.7)	1.0 (reference)				
			GA	13 (18.6)					
			AA	53 (75.7)					
			A allele carrier	66 (94.3)	1.55 (0.20-11.54)	0.44	.67		
CTLA-4	2q33.2	rs231775	AA	34 (48.6)	1.0 (reference)				
			AG	26 (37.1)					
			GG	10 (14.3)					
			G allele carrier	36 (51.4)	0.22 (0.25-0.99)	-1.48	.04	2.091 (0.233-2.219)	.05

[a]Univariate analysis was carried out without any adjustment in order to generate hazard ratios with confidence intervals for individual risk for each of the parameters on survival.

Table 5. Univariate and multivariate cox proportional hazards models of associations of overall survival with clinicopathologic characteristics and immune cell profiles

	Univariate Analysis[a]			Multivariate Analysis[b]	
	HR[c] (95% CI)	**HR[c]**	***P***	**HR[c] (95% CI)**	***P***
Clinicopathological Characteristics					
Age (yrs): >65 vs ≤65	0.99 (0.97-1.01)	-0.005	.65		
Gender					
Female	2.18 (1.29-3.45)	0.75	.003		
History of Tobacco					
Yes	1.01 (1.009-1.016)	0.009	.007		
Tumor Size (cm)					
>4.5	1.23 (1.11-1.37)	0.21	<.001		
Lymph Node Status (N)					
N_0	0.387 (0.219-0.684)	-0.95	<.001		
N_1 *(reference)*					
Recurrence-free disease (RFD)	0.77 (0.47-1.25)	-0.26	.29		
Stage					
I	0.369 (0.225-0.605)	-0.997	<.001	0.437	.006
>II (reference)				(0.243-0.785)	
Histologic Types					
ADC	2.782 (0.857-9.037)	1.023	.08		
SCC	3.231	1.173	.05	2.989	.02
LCC (reference)	(0.971-10.747)			(1.123-7.954)	
Adjuvant Therapy					
Yes	1.724 (1.038-2.864)	-0.545	.03	0.428	.02
				(0.200-0.914)	
Immune Cells (≤ median *vs* > median)					
MCs PD-L1+/mm^2					
>1.03	1.015 (0.589-1.750)	0.015	.95		
CD3+ cells/mm^2					
>327.33	1.077 (0.625-1.857)	0.075	.78		
CD3+CD8+ cells/mm^2					
>110.35	1.744 (1.003-3.031)	0.556	.04		
CD3+PD-1+ cells/mm^2					
>40.36	1.264 (0.731-2.186)	0.234	.40		
CD68+ cells/ mm^2					
>77.81	1.125 (0.653-1.939)	0.118	.67		
CD68+PD-L1 cells/mm^2					
>2.54	1.234 (0.715-2.129)	0.210	.45		
CD3+CD45RO+ cells/mm^2					
> 562.45	1.716 (0.984-2.993)	0.540	.05	2.265 (1.276-4.022)	.005
CD3+CD45RO+FOXP3+ cells/mm^2					
>126.84	1.923 (1.090-3.396)	0.654	.02		
CD3+CD57+ cells/mm^2					
>22.07	1.085 (0.624-1.884)	0.081	.77		
CD20+ cells/mm^2					
>29.54	1.306 (0.753-2.263)	0.267	.34		
EMT gene mutations (% of samples)					
CD44 (2.8)	0.899 (0.666–1.208)	0.573	.44		
CDH1 (2.8)	1.414 (0.082–2.357)	0.345	.80		
MMP2 (5.7)	0.091 (0.022–0.459)	-2.409	.64		
ZEB1 (2.8)	1.166 (0.155–8.659)	0.147	.88		
ZEB2 (2.8)	0.151 (0.011–1.207)	-1.891	.04		.04

	1.429 (0.240–1.972)

Abbreviations: EMT, epithelial-mesenchymal transition; HR, hazard ratio (b coefficient); LCC, large cell carcinoma; MC, malignant cell; PD-L1, programmed death ligand-1; SCC, squamous cell carcinoma.
[a] Univariate analysis was carried out without any adjustments to generate hazard ratios for overall survival with confidence intervals for each parameter.
[b] Multivariate analysis was carried out to analyze the effects of several risk parameters on overall survival.

DISCUSSION

In the present study, we explored the quantitative relationships among TAIC density and EMT and immune checkpoint genomic profiles in 164 patients with nonmetastatic NSCLC. Our results indicate that incorporating genomic profile data into quantitative immune profiling produced several important findings. First, our results validate mIF for immune profiling in NSCLC. Second, we discovered associations between clinicopathologic characteristics of NSCLC and PD-L1 expression by MCs and immune cells. Third, our analyses revealed that higher densities of antigen-experienced T cells were associated with metastasis. Fourth, we found that microenvironment types II (immunologic ignorance) and IV (tolerance) were the most frequent microenvironments in the tested NSCLC tissues. Fifth, we found that patients with early-stage disease, carriers of the G allele of *CTLA4* variant rs231775, patients with wild-type *ZEB2*, and patients with a high density of CD3+CD45RO+ T cells had longer OS than other patients. Finally, we found that ZEB2 mutations were indicative of EMT and served as a proxy for noninflammatory immune phenotypes (immunologic ignorance and immune tolerance). This information may be important for predicting response to immune checkpoint blockade, particularly in clinical settings where NGS is available but immune profiling is not.

To better understand the role of immune and genomic profiling in early-stage NSCLC, our study was carried out in a cohort of patients with nonmetastatic disease. Our investigation included adenocarcinoma, SCC, and LCC, and, to our knowledge, it is the first study to compare

genetic and immune cell profiles among the 3 major histologic types of NSCLC and to determine their impact on patient outcomes.

Using mIF and image analysis, we found that PD-L1 expression by MCs and other cells was associated with several clinicopathologic characteristics and with disease outcomes. We found no differences in PD-L1 expression on MCs among the histologic tumor types, but our results showing that 31.51% of LCC samples had positive PD-L1 expression on MCs agreed with those of several studies of large cell neuroendocrine carcinoma of the lung.[22,23] This similarity suggests that patients with the large cell subtype of NSCLC may benefit from checkpoint inhibitor therapies. In agreement with findings reported by Calles and colleagues,[24] we found that high expression of PD-L1 on MCs was associated with tobacco use and the adenocarcinoma histologic type. Interestingly, we found that low densities of natural killer T cells were also associated with tobacco use; this finding was in concordance with Hogan and colleagues'[25] study, which showed a significant reduction of natural killer cells in tobacco users. That reduction was accompanied by significant defects in cytokine production, suggesting that depletion of natural killer T cells can interfere with antitumor immune responses.

Similarly, our study showed that low densities of cytotoxic T cells were associated with LCC and that low densities of memory T cells were associated with female sex and older age, suggesting that alterations in the regulation of the immune system can increase the likelihood of tumor invasion and progression. As suggested by Prado-Garcia and colleagues,[26] alterations in the immune system induced by tumor cells can lead to T-cell dysfunction and reduce T cells' ability to attack tumors. It is also known that increasing age can produce downregulation of immune cells.[27]

Interestingly, shorter survival time was associated with high densities of CD3+PD-1+ antigen–experienced T cells, suggesting that PD-1 plays an important role in facilitating NSCLC metastasis. A similar finding was made by Wang et al,[28] who found that, in cervical cancer, the

interaction between PD-1 and PD-L1 can trigger immune responses and facilitate tumor growth and metastasis. Most, if not all, malignancies trigger different immune responses through intricate interactions between tumor cells and the host's immune cells,[29] suggesting that the progression of cancer is influenced by complex tumor–3immune cell interactions driven by characteristics of the tumor, the behavior of inflammatory cells, and genetic mutations.

Our study also identified differences in the immune profiles of the histologic types of NSCLC that may affect the choice of therapy. Densities of different subpopulations of T cells, including cytotoxic T cells, were higher in adenocarcinoma and LCC than in SCC; these results agreed with those of our previous study.[30] In addition, we also observed higher densities of CD3+PD-1+ antigen–experienced T cells in adenocarcinoma and LCC than in SCC, suggesting that adenocarcinoma and LCC provoke more immune exhaustion and, thus, that immunotherapeutic intervention may be warranted for these 2 histologic types. Previous studies have shown that activation of the PD-1 pathway mediates inhibition of T cells and that PD-1 expression is an independent predictive factor for prognosis.[31] Using the tumor microenvironment criteria proposed by Teng and colleagues,[20] we determined that the most common microenvironments in our cohort were immunologic ignorance (PD-L1/TIL) and immunologic tolerance (PD-L1/ TILþ), suggesting that complementary strategies, as proposed by Teng et al,[20] will be necessary to induce a response to immunotherapy in patients with these microenvironments.

Our findings also highlighted the role of EMT in genomic profiling to predict outcomes in NSCLC. The EMT is activated in cancer cells and is induced by transcription factors such as *Snail*, *Twist*, and *ZEB1/ZEB*2 to modify the transcriptional machinery, alter translation and protein stability, and promote invasion and metastasis.[32] In fact, we demonstrated that *ZEB1* mutations were inversely correlated with levels of CD3+ and CD3+CD8+ T cells, suggesting a molecular link between EMT and immunosuppression, 2 key drivers of cancer progression. We

also found that mutations in the EMT-associated gene *CD44* were associated with adenocarcinoma and SCC and with tobacco use, in agreement with other studies.[32–34] Interestingly, tobacco use induces metaplastic bronchial squamous epithelium that exhibits increased hyaluronan and CD44 expression in the proliferating basal layers. In premalignant bronchial dysplasia, the entire epithelial thickness shows aberrant hyaluronan-CD44 expression, indicating that squamous malignant transformation is closely associated with CD44 expression and with tobacco use.[35] We also found that CD44-mutant tumors were associated with high densities of CD20+ B cells, which suggests that CD44 also may participate in several other immune-related processes, as previously demonstrated by Zittermann and colleagues.[36]

Over the last decade, significant progress has been made in systemic therapies to improve the length of good-quality survival in patients with metastatic NSCLC. However, many questions remain about the behavior of nonmetastatic NSCLC. One such issue relates to the expected 5-year survival rate, which ranges from 36% to 92% for patients with early-stage (ie, stage I, II, or IIIA) disease.[3] Another question relates to current studies of immunotherapy in the context of nonmetastatic NSCLC.3 Not all NSCLC patients have a response to immunotherapy because the heterogeneity of immune cell profiles in NSCLC tumors probably depends on other factors, for instance EMT and immune checkpoint genes' control of cancer progression.[7,37] In this context, in our cohort, multivariate analysis reliably predicted longer OS for patients with a high density of CD3+CD45RO+ memory T cells and those with the G allele of *CTLA4* variant rs231775 and shorter OS when nonmetastatic NSCLC tumors have *ZEB2* mutations. In fact, the presence of a high density of CD3+CD45RO+ immune cells in the tumor region is correlated with favorable clinical outcomes in epithelial lung cancers.[38] In addition, several studies have assessed the contribution of *CTLA4* polymorphisms in the context of immune checkpoint inhibition.[39–44] Our multivariate Cox analysis revealed that the *CTLA4* rs231775 AA genotype is an adverse prognostic indicator for NSCLC patients. Although associations between *CTLA4* and outcomes

have been reported, these lacked independent validation. As shown by Deng and colleagues,[44] CTLA4 expression was not significantly prognostic in their cohort of 1432 patients on univariate analysis but was significant in their cohort of patients with available data for multivariate analysis.

Our analysis has a number of limitations that could not be addressed in this first quantitative study. The first of these limitations is the small cohort, which included 164 retrospectively collected cases. However, the maximum follow-up time was 168 months, the robustness of clinical variables was ascertained, and survival information was available. A second limitation is that we included only 70 tissue samples in our genomic analysis. However, a preliminary examination of Kaplan-Meier survival curves (data not shown) demonstrated that patients with pathologic stages II and III disease had approximately the same hazard for survival, with a median survival time of 40 months for both groups. A third limitation of the study concerns the noninclusion of normal tissue or blood samples for the genomic analysis, which may have affected the allele frequency determination. A fourth limitation was the tissue microarray format we used, which could have induced underrepresentation or overrepresentation of PDL1 due to tumor heterogeneity and the small area of analysis. Fifth, we lacked available data regarding responses to anti–PD-1/anti–PD-L1 monoclonal antibodies in our patient population. Sixth, even with accumulating evidence supporting the contribution of germline genetics to host immunity, knowledge about host genetic factors as predictive biomarkers of clinical outcomes is limited. Moreover, at present, no systematic study of genetic variants as surrogates for immunotherapy outcomes at the genome-wide level is available. Larger study cohorts will be required for the discovery of low-penetrance germline loci associated with immunotherapy outcomes. This clearly indicates the need for a large international collaboration pooling patient resources. The implementation of genome-wide association studies coupled with large-scale immune phenotyping will help to identify common genetic variations associated with a large population–

based scale.[42] Finally, mIF can be used to simultaneously identify specific proteins on a single slide. This technique allows study of the immune contexture using paraffin-embedded tumor tissues. In addition, multiplexed image analysis methods are highly advantageous for investigating immune evasion mechanisms and for discovering potential biomarkers to assess mechanisms of action and to predict response to a given treatment.[45,46] The combination of NGS data with mIF may facilitate disease management and the appropriate application of immunotherapeutic agents.

In conclusion, incorporating a genomic profile into quantitative mIF revealed a variety of factors associated with the behavior of nonmetastatic NSCLC, including specific clinicopathologic characteristics and immunomodulatory control of local disease progression. Thus, these tools might be useful for predicting whether patients with early-stage, nonmetastatic NSCLC will benefit from combinations of targeted therapy, chemotherapy, and immunotherapy

References

1.Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. J Thorac Oncol. 2016;11(1):39-51.

2.Ghysen K, Vansteenkiste J. Immunotherapy in patients with early stage resectable nonsmall cell lung cancer. Curr Opin Oncol. 2019;31(1):13-17.

3.Deslypere G, Gullentops D, Wauters E, Vansteenkiste J. Immunotherapy in non-metastatic non-small cell lung cancer: Can the benefits of stage IV therapy be translated into earlier stages? Ther Adv Med Oncol. 2018;10:1758835918772810.

4.Yu H, Boyle TA, Zhou C, et al. PD-L1 expression in lung cancer. J Thorac Oncol. 2016;11:964-975.

5.Sun JM, Zhou W, Choi YL, et al. Prognostic significance of PD-L1 in patients with non-small cell lung cancer: a large cohort study of surgically resected cases. J Thorac Oncol. 2016;11(7):1003–1011.

6 .Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. N Engl J Med. 2016;375:1823-1833.

7.Kim S, Koh J, Kim MY, et al. PD-L1 expression is associated with epithelial-to-mesenchymal transition in adenocarcinoma of the lung. Hum Pathol. 2016;58:7-14.

8.Marwitz S, Scheufele S, Perner S, Reck M, Ammerpohl O, Goldmann T. Epigenetic modifications of the immune-checkpoint genes CTLA4 and PDCD1 in non-small cell lung cancer results in increased expression. Clin Epigenetics. 2017;9:51-53.

9.Papanicolau-Sengos A, Yang Y, Pabla S, et al. Identification of targets for prostate cancer immunotherapy. Prostate. 2019;79:498-505.

10.O'Leary K, Shia A, Schmid P. Epigenetic Regulation of EMT in Non-Small Cell Lung Cancer. Curr Cancer Drug Targets. 2018;18:89-96.

11.Chae YK, Chang S, Ko T, et al. Epithelial-mesenchymal transition (EMT) signature is inversely associated with T-cell infiltration in non-small cell lung cancer (NSCLC). Sci Rep. 2018;8:2918-2926.

12.Chen L, Gibbons DL, Goswami S, et al. Metastasis is regulated via microRNA-200/ZEB1 axis control of tumour cell PD-L1 expression and intratumoral immunosuppression. Nat Commun. 2014;5:5241-5257.

13.Kurimoto R, Ebata T, Iwasawa S, et al. Pirfenidone may revert the epithelial-to-mesenchymal transition in human lung adenocarcinoma. Oncol Lett. 2017;14:944-950.

14.Alsuliman A, Colak D, Al-Harazi O, et al. Bidirectional crosstalk between PD-L1 expression and epithelial to mesenchymal transition: significance in claudin-low breast cancer cells. Mol Cancer. 2015;14:149-162.

15.Ma Y, Adjemian S, Mattarollo SR, et al. Anticancer chemotherapy-induced intratumoral recruitment and differentiation of antigen-presenting cells. Immunity. 2013;38:729-741.

16.Mazzaschi G, Madeddu D, Falco A, et al. Low PD-1 Expression in Cytotoxic CD8(+) Tumor-Infiltrating Lymphocytes Confers an Immune-Privileged Tissue Microenvironment in NSCLC with a Prognostic and Predictive Value. Clin Cancer Res. 2018;24:407-419.

17.Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of Lung Tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol. 2015;10(9): 1243–1260.

18. Parra ER, Francisco-Cruz A, Wistuba, II. State-of-the-Art of Profiling Immune Contexture in the Era of Multiplexed Staining and Digital Analysis to Study Paraffin Tumor Tissues. Cancers (Basel). 2019;11. pii: E247.

19. Gorris MAJ, Halilovic A, Rabold K, et al. Eight-color multiplex immunohistochemistry for simultaneous detection of multiple immune checkpoint molecules within the tumor microenvironment. J Immunol. 2018;200(1): 347–354.
20.Teng MW, Ngiow SF, Ribas A, Smyth MJ. Classifying cancers based on T-cell infiltration and PD-L1. Cancer Res. 2015;75(11):2139–2145.
21.Machado-Rugolo J, Fabro AT, Ascheri D, et al. Usefulness of complementary next-generation sequencing and quantitative immunohistochemistry panels for predicting brain metastases and selecting a treatment outcomes of non-small cell lung cancer. Hum Pathol. 2019;83:177–191.
22. Eichhorn F, Harms A, Warth A, Muley T, Winter H, Eichhorn ME. PD-L1 expression in large cell neuroendocrine carcinoma of the lung. Lung Cancer. 2018;118:76–82.
23.Tsuruoka K, Horinouchi H, Goto Y, et al. PD-L1 expression in neuroendocrine tumors of the lung. Lung Cancer. 2017;108:115–120.
24. Calles A, Liao X, Sholl LM, et al. Expression of PD-1 and its ligands, PD-L1 and PD-L2, in smokers and never smokers with KRAS-mutant lung cancer. J Thorac Oncol. 2015;10(12):1726–1735.
25. Hogan AE, Corrigan MA, O'Reilly V, et al. Cigarette smoke alters the invariant natural killer T cell function and may inhibit anti-tumor responses. Clin Immunol. 2011;140(3):229–235.
26.Prado-Garcia H, Romero-Garcia S, Aguilar-Cazares D, Meneses-Flores M, Lopez-Gonzalez JS. Tumor-induced CD8þ T-cell dysfunction in lung cancer patients. Clin Dev Immunol. 2012;2012:741741.
27. Fulop T, Larbi A, Kotb R, de Angelis F, Pawelec G. Aging, immunity, and cancer. Discov Med. 2011;11(61):537–550.
28. Wang S, Li J, Xie J, et al. Programmed death ligand 1 promotes lymph node metastasis and glucose metabolism in cervical cancer by activating integrin beta4/SNAI1/SIRT3 signaling pathway. Oncogene. 2018;37(30):4164–4180.
29. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nature. 2008;454(7203):436–444.
30. Parra ER, Behrens C, Rodriguez-Canales J, et al. Image analysis-based assessment of PD-L1 and tumor-associated immune cells density supports distinct intratumoral microenvironment groups in non-small cell lung carcinoma patients. Clin Cancer Res. 2016;22(24):6278–6289.
31. Paulsen EE, Kilvaer TK, Khanehkenari MR, et al. Assessing PDL-1 and PD-1 in non-small cell lung cancer: a novel immunoscore approach. Clin Lung Cancer. 2017;18(2):220–233 e8.
32. Gonzalez DM, Medici D. Signaling mechanisms of the epithelialmesenchymal transition. Sci Signal. 2014;7(344): re8.

33. Leung EL, Fiscus RR, Tung JW, et al. Non-small cell lung cancer cells expressing CD44 are enriched for stem cell-like properties. PLoS One. 2010; 5(11):e14062

34. Chou YE, Hsieh MJ, Chiou HL, Lee HL, Yang SF, Chen TY. CD44 gene polymorphisms on hepatocellular carcinoma susceptibility and clinicopathologic features. Biomed Res Int. 2014;2014:231474.

35. Penno MB, August JT, Baylin SB, et al. Expression of CD44 in human lung tumors. Cancer Res. 1994;54(5):1381–1387.

36. Zittermann SI, Achino BI, Agriello EE, Halperin N, Ramhorst RE, Fainboim L. Modulation of CD44 in acute lymphoblastic leukemia identifies functional and phenotypic differences of human B cell precursors. Eur J Haematol. 2001;66(6): 377–382.

37. Koh J, Go H, Keam B, et al. Clinicopathologic analysis of programmed cell death-1 and programmed cell death-ligand 1 and 2 expressions in pulmonary adenocarcinoma: comparison with histology and driver oncogenic alteration status. Mod Pathol. 2015;28(9):1154–1166.

38. Sheng SY, Gu Y, Lu CG, Zou JY, Hong H, Wang R. The distribution and function of human memory T cell subsets in lung cancer. Immunol Res. 2017; 65(3):639–650.

39. Breunis WB, Tarazona-Santos E, Chen R, Kiley M, Rosenberg SA, Chanock SJ. Influence of cytotoxic T lymphocyte-associated antigen 4 (CTLA4) common polymorphisms on outcome in treatment of melanoma patients with CTLA-4 blockade. J Immunother. 2008;31(6):586–590.

40. Queirolo P, Morabito A, Laurent S, et al. Association of CTLA-4 polymorphisms with improved overall survival in melanoma patients treated with CTLA-4 blockade: a pilot study. Cancer Invest. 2013;31(5):336–345.

41. Hamid O, Schmidt H, Nissan A, et al. A prospective phase II trial exploring the association between tumor microenvironment biomarkers and clinical activity of ipilimumab in advanced melanoma. J Transl Med. 2011;9:204–220.

42. Kirchhoff T, Ferguson R. Germline genetics in immuno-oncology: from genome-wide to targeted biomarker strategies. Methods Mol Biol. 2020;2055:93– 117.

43. Lim YW, Chen-Harris H, Mayba O, et al. Germline genetic polymorphisms influence tumor gene expression and immune cell infiltration. Proc Natl Acad Sci USA. 2018;115:E11701–E11710.

44. Deng L, Gyorffy B, Na F, et al. Association of PDCD1 and CTLA-4 gene expression with clinicopathological factors and survival in non-small-cell lung cancer: results from a large and pooled microarray database. J Thorac Oncol. 2015;10(7):1020–1026.

45. Hofman P, Badoual C, Henderson F, et al. Multiplexed immunohistochemistry for molecular and immune profiling in lung cancer—just about ready for prime-time? Cancers (Basel). 2019;11(3):283.

46. Parra ER, Francisco-Cruz A, Wistuba II. State-of-the-art of profiling immune contexture in the era of multiplexed staining and digital analysis to study paraffin tumor tissues. Cancers (Basel). 2019;11(2):247.

Supplementary Table 1. Target genes in the TruSeq Custom Amplicon Panel.

Pathway	Gene	Function	Target region	Number of targets (N = 121)	Number of amplicons (N = 372)
Immune checkpoints	*CD276*	Regulation of T-cell-mediated immune response	Chr 15: 73683966-73714518	8	21
	CTLA4	Downregulation of immune responses	Chr 2: 203867788-203873960	4	12
	CD274 (PD-L1)	Regulation of T-cell-mediated immune response Upregulation of T-cell-mediated immune response by IFNG/IFN	Chr 9:5450381-5470567	7	27
	PDCD1LG2 (PD-L2)	gamma stimulation	Chr 9: 5510438-5571282	7	18
	LAG3	DNA repair	Chr 12: 6772483-6778455	6	17
	VTCN1 (B7H4)	Negative modulation in cellular immune response	Chr 1: 117143587-117270368	7	21
Epithelial-to-mesenchymal transition genes	*CD44*	Cell–cell interactions, cell adhesion, and migration	Chr 11: 35138870-35232402	19	48
	CDH1 (E-cadherin)	Cellular adhesion	Chr 16: 68737290-68835542	14	32
	TGFβ1	Notch signaling	Chr 19: 41330323-41353933	6	15
	VIM	Transcription factor	Chr 10: 17226969-17238563	9	16
	ZEB1	Transcription factor	Chr 10: 31318495-31529814	10	44
	MMP2	Cellular adhesion	Chr 16: 55478830-55506691	13	25
	ZEB2	Transcription factor	Chr 2: 144384375-144520393	11	76

Supplementary Table 2. Associations between clinicopathologic characteristics, immune cell profiles, genetic variants, and somatic alterations.

	Clinicopathologic Parameters															
	Age (y)		**Sex**		**Tobacco use**		**Lymph node status**		**Brain metastasis**		**Stage**		**Histologic type**		**Adjuvant treatment**	
	<65 *vs* ≥65		**Male *vs* Female**		**No *vs* Yes**		**N0 *vs* N1 and N2**		**No *vs* Yes**		**I *vs* II and III**		**SCC *vs* ADC *and* LCC**		**No *vs* Yes**	
Immune profile	β	*P*	β	*P*	β	*P*	β	*P*	β	*P*	β	*P*	β	*P*	β	*P*
MCs PD-L1+	-.33	.73	1.34	.28	2.85	.04	-.36	.92	.20	.96	.61	.84	2.85	.04	.29	.82
CD3+	-209.81	.09	-196.94	.36	-57.35	.83	-538.14	.40	-61.61	.47	314.41	.56	-94.82	.80	1.31	.96
CD3+CD8+	-83.29	.08	-52.17	.53	-98.05	.37	-359.57	.15	-29.14	.39	402.54	.05	-312.87	.04	-77.10	.38
CD3+PD1+	-32.67	.23	-57.98	.16	53.79	.32	-25.83	.83	-19.79	.04	19.52	.85	6.30	.41	19.52	.65
CD68+	-58.35	.25	86.31	.35	222.11	.05	47.86	.86	99.08	.78	22.18	.92	-18.64	.27	64.94	.50
CD68+PD-L1+	-3.27	.20	41.50	.28	68.24	.18	34.80	.72	-22.90	.88	34.80	.72	-41.42	.55	3.75	.45
CD3+CD45RO+	-216.95	.04	-441.56	.01	415.03	.05	24.71	.96	-695.58	.33	194.24	.66	145.87	.66	-18.94	.30
CD3+CD57+	-75.24	.75	214.60	.58	-1073.75	.04	-164.19	.88	1326.89	.40	413.01	.67	157.93	.82	89.15	.82
CD3+CD45RO+FOXP3+	-109.51	.12	-149.64	.24	413.59	.01	-103.93	.78	-26.01	.61	357.83	.26	-209.20	.35	-224.94	.05
CD20+	76.90	.49	-13.37	.33	148.24	.41	-495.96	.22	-308.57	.57	56.86	.09	17.47	.47	144.92	.04
SNVs in immune genes	χ^2	*P*	χ^2	*P*	χ^2	*P*	χ^2	*P*	χ^2	*P*	χ^2	*P*	χ^2	*P*	χ^2	*P*
CD274 rs2297136 GA/AA vs. GG	1.59	.65	4.12	.11	5.49	.12	4.13	.63	1.80	.33	2.21	.75	2.39	.72	3.32	.29
CTLA-4 rs231775 AG/GG vs. AA	3.78	.15	1.89	.38	.25	.88	.74	.94	.16	.92	11.36	.02	4.78	.31	7.82	.25
Somatic alterations in the EMT genes	χ^2	*P*	χ^2	*P*	χ^2	*P*	χ^2	*P*	χ^2	*P*	χ^2	*P*	χ^2	*P*	χ^2	*P*
CD44	3.40	.76	6.49	.37	14.18	.02	15.45	.21	6.34	.38	9.55	.65	11.68	.47	22.09	.23
CDH1	3.44	.33	2.88	.41	5.43	.14	2.09	.91	.46	.93	4.52	.60	4.58	.59	1.35	.71
MMP2	.27	.60	.83	.36	1.49	.22	1.70	.42	.56	.45	.86	.65	.19	.90	.84	.83
ZEB1	3.42	.18	.84	.65	4.83	.05	1.95	.74	.27	.87	1.57	.81	5.12	.26	1.35	.51
ZEB2	3.82	.43	3.93	.41	1.64	.65	4.30	.82	1.12	.89	4.61	.79	2.89	.94	7.63	.10

SCC, squamous cell carcinoma; ADC, adenocarcinoma; LCC, large cell carcinoma; SNV, single-nucleotide variant; EMT, epithelial-mesenchymal transition.
[a]Multivariate analysis was carried out to analyze the effect of several dependent variables on clinicopathologic parameters.
The bolded values were statistically significant.

Supplementary Table 3. Gene mutations found in 70 primary non-small cell lung carcinomas from the study cohort and 230 lung adenocarcinoma samplesfrom The Cancer Genome Atlas (TCGA) database.

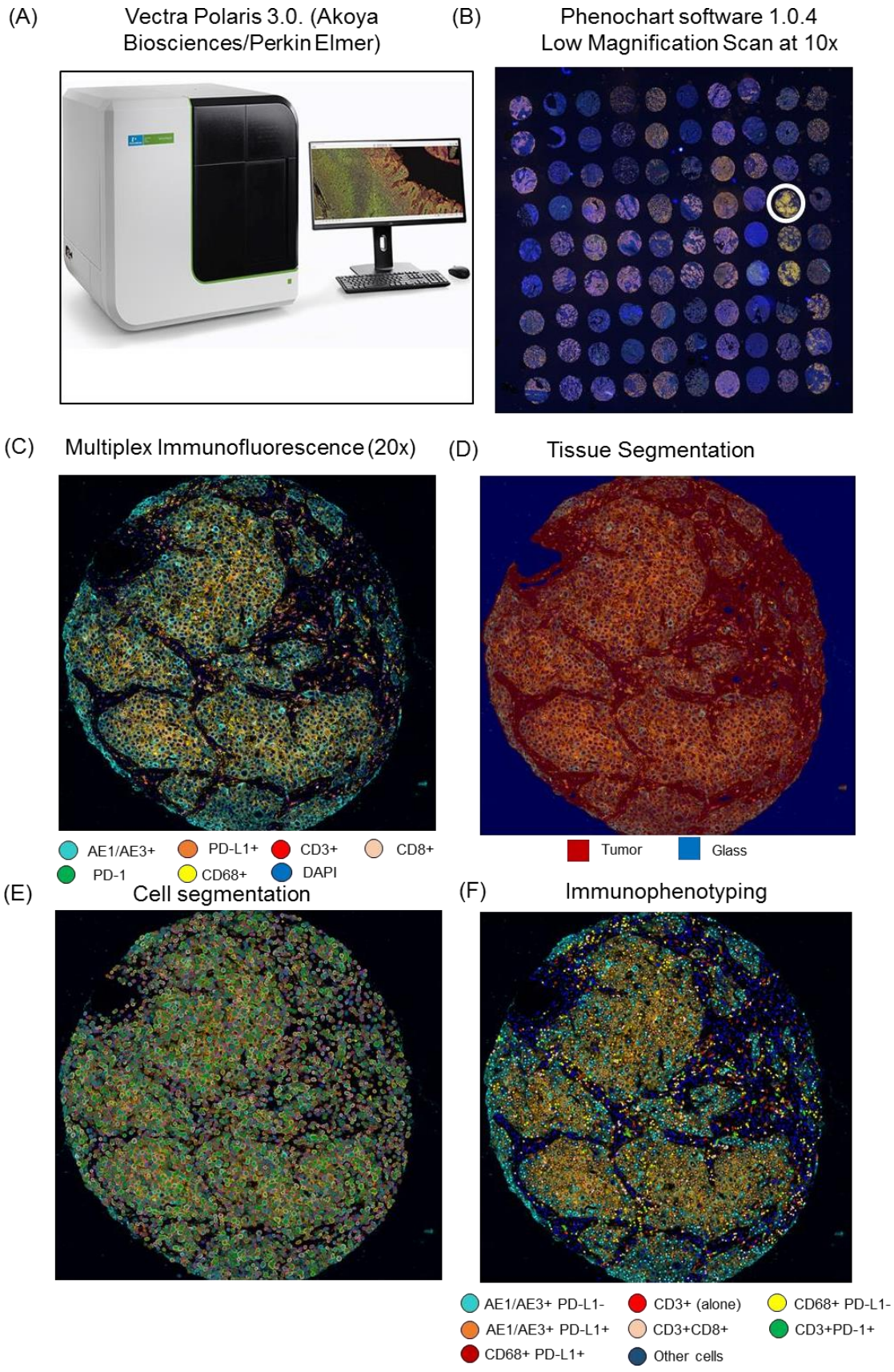
Function	Gene	Study cohort (N = 70)	TCGA data[a] (N = 230)
Immune checkpoints	*CD276*	1 (1%)	1 (0.4%)
	CTLA-4	0	3 (1.3%)
	CD274 (PD-L1)	0	2 (0.9%)
	PDCD1LG2 (PD-L2)	0	1 (0.4%)
	LAG3	0	0
	VTCN1 (B7H4)	0	2 (0.9%)
Epithelial-mesenchymal transition genes	*CD44*	2 (3%)	1 (0.4%)
	CDH1 (E-cadherin)	2 (3%)	3 (1.3%)
	TGFβ1	1 (1%)	2 (0.9%)
	VIM	1 (1%)	3 (1.3%)
	ZEB1	2 (3%)	13 (5.2%)
	MMP2	4 (6%)	12 (4.8%)
	ZEB2	10 (14%)	17 (7.4%)

[a] TCGA data were obtained from cBioPortal for Cancer Genomics (https://cbioportal.org).

Supplementary Table 4. Univariate analysis of correlation between PD-L1 expression, tumor-associated immune cell levels, and genetic variantsin non-small cell lung cancer.

Marker	Variant	R^2	*P*
	SNV		
CD3+PD1+ (cell/mm^2)	*CTLA-4* rs231775 GG genotype	.33	.01
	Somatic alteration		
CD3+ (cell/mm^2)	*ZEB1*	.27	.04
CD3+CD8+ (cell/mm^2)	*ZEB1*	.34	.01
CD3+PD1+ (cell/mm^2)	*CD276*	.30	.02
CD20+ (cell/mm^2)	*CD44*	.29	.03

Supplementary Figure 1. Representative multiplex immunofluorescence workflow showing differentsteps in the process from scanning to analysis of the images.



A) Vectra Polaris 3.0 multispectral microscope used for the analysis. (B) Tissue microarray (TMA) image scanned at ×10 magnification. For further analysis, samples were scanned at high magnification (20x), core by core. (C) Multiplex immunofluorescence image scanned at 20x magnification and viewedin InForm 2.1.3 image analysis software. (D) Tissue segmentation (tumor and glass) using InForm software. (E) Cell segmentation to characterize individual cells; 4′,6-diamidino-2-phenylindole (DAPI) was used as a counterstaining marker. (E) Cell immunophenotyping, used to identify subgroups of cells,was supervised by a pathologist.

5.2 Artigo 2: citação completa (ANEXO 2)

Machado-Rugolo J, Gutierrez Prieto T, Fabro AT, Parra Cuentas ER, Sá VK, Baldavira CM, Rainho CA, Castelli EC, Farhat C, Takagaki TY, Nagai MA, Capelozzi VL. Relevance of PD-L1 Non-Coding Polymorphisms on the Prognosis of a Genetically Admixed NSCLC Cohort. Pharmgenomics Pers Med. 2021;14:239-252

doi: https://doi.org/10.2147/PGPM.S286717

Fator de impacto = 2.606

Relevance of PD-L1 non-coding polymorphisms on the prognosis of a genetically admixed NSCLC cohort

Juliana Machado-Rugolo[1,2], Tabatha Gutierrez Prieto[1], Alexandre Todorovic Fabro[3], Edwin Roger Parras Cuentas[4], Vanessa Karen Sá[5], Camila Machado Baldavira[1] Claudia Aparecida Rainho[6], Erick C. Castelli[7,8], Cecilia Farhat[1], Teresa Yae Takagaki[9], Maria Aparecida Nagai[10,11], Vera Luiza Capelozzi[1*]

[1]Laboratory of Genomics and Histomorphometry, Department of Pathology, University of São Paulo Medical School (USP), São Paulo, Brazil;
[2]Health Technology Assessment Center, Clinical Hospital (HCFMB), Medical School of São Paulo State University (UNESP), Botucatu, São Paulo, Brazil;
[3]Department of Pathology and Legal Medicine, Ribeirão Preto School of Medicine, University of São Paulo (FMRP-USP), Ribeirão Preto, Brazil;
[4]Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA;
[5]Laboratory of Genomics and Molecular Biology, Centro Internacional de Pesquisa (CIPE), AC Camargo Cancer Center, São Paulo, SP, Brazil;
[6]Department of Genetics, Biosciences Institute, São Paulo State University (UNESP), Botucatu, Brazil;
[7]Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unit (UNIPEX), Medical School of São Paulo State University (UNESP), Botucatu, São Paulo, Brazil;
[8]Department of Pathology, Medical School of São Paulo State University (UNESP), Botucatu, São Paulo, Brazil;
[9]Division of Pneumology, Instituto do Coração (Incor), Faculty of Medicine, University of São Paulo, São Paulo, Brazil;
[10]Department of Radiology and Oncology, University of São Paulo Medical School (USP), São Paulo, Brazil;
[11]Laboratory of Molecular Genetics, Center for Translational Research in Oncology, Cancer Institute of São Paulo (ICESP), São Paulo, Brazil.

Abstract:

Purpose: Although non-small cell lung cancer (NSCLC) remains a deadly disease, new predictive biomarkers have emerged to assist in managing the disease, of which one of the most promising is the programmed death-ligand 1 (PD-L1). Each, *PD-L1* variant seem to modulate the function of immune checkpoints differently and affect response to adjuvant treatment and outcome in NSCLC patients. We thus investigated the influence of these PD-L1 genetic variations in genetically admixed NSCLC tissue samples, and correlated these values with clinicopathological characteristics, including prognosis.

Patients and methods: We evaluated PD-L1 non-coding genetic variants and protein expression in lung adenocarcinomas (ADC), squamous cell carcinomas (SqCC), and large cell carcinomas (LCC) *in silico*. Microarray paraffin blocks from 70 samples of ADC (N=33), SqCC (N=24), and LCC (N=13) were used to create PD-L1 multiplex immunofluorescence assays with a Cell Signaling E1L3N clone. Fifteen polymorphisms of the *PD-L1* gene were investigated by targeted sequencing and evaluated *in silico* using dedicated tools.

Results: Although *PD-L1* polymorphisms seemed not to interfere with protein expression, PD-L1 expression varied among different histological subtypes, as did clinical outcomes, with the rs4742098A>G, rs4143815G>C, and rs7041009G>A variants being associated with relapse (P=0.01; P=0.05; P=0.02, respectively). The rs7041009 GG genotype showed a significant correlation with younger and alive patients compared to carriers of the A allele (P=0.02 and P<0.01, respectively). The Cox regression model showed that the rs7041009 GG genotype may influence OS (P<0.01) as a co-dependent factor associated with radiotherapy and recurrence in NSCLC patients. Furthermore, the Kaplan-Meier survival curves showed that rs7041009 and rs4742098 might impact PPS in relapsed patients. *In silico* approaches identified the variants as benign.

Conclusion: *PD-L1* non-coding variants play an important role in modulating immune checkpoint function and may be explored as immunotherapy biomarkers. We highlight the rs7041009 variant, which impacts OS and PPS in NSCLC patients.

Keywords: *PD-L1*, Non-small Cell Lung Cancer, Single Nucleotide Polymorphisms, Next-Generation Sequencing

Introduction

Lung cancer is one of the leading causes of cancer-related death globally.[1] Although there are different types of lung cancer, non-small cell lung cancer (NSCLC) represents 85% of all primary lung tumors. NSCLC is a grim disease that is aggravated by the fact that patients normally either receive their diagnosis at advanced stages or present with recurrent disease after initial locoregional treatment.[2] Over the last few decades, conventional chemotherapy, mainly platinum-based chemotherapy, used to be the only therapeutic option for those not eligible for radical intent treatment: a treatment with limited efficacy and very few long-term survivors (5-year overall survival less than 15%). Furthermore, these patients often lacked therapeutic options beyond first-line treatment.[3]

More recently, though, immune checkpoint molecules involved in tumor immune evasion were identified and immune checkpoint inhibitors (ICIs) were introduced in antitumor immunotherapy. This new therapeutic approach targets an inhibitory receptor, the programmed cell death-1 (PD-1) receptor, to assist the immune system in identifying and neutralizing malignant cells. However, tumor cells may evade the host immunosurveillance by expressing the programmed death-1-ligand 1 (PD-L1) as an adaptive, resistant mechanism to suppress this inhibitory receptor.[4] Thus, because PD-L1 up-regulation by tumor cells can protect them from antitumor immune response, the blockade of PD-L1/ PD-1 interactions has been recently selected for antitumor immune therapy.[5] Agents targeting the PD-1/PD-L1 signaling pathway have shown promising responses in different types of cancer, including NSCLC. These results point to PD-L1 protein expression as a potential predictive marker for a successful blockade of PD-L1/PD-1 interactions.[6] However, several challenges remain in producing robust evidence to support the use of this biomarker.

In this context, some studies with NSCLC patients have demonstrated that those with more than 50% PD-L1 positive tumor cells are non-responders to anti-PD-1/PD-L1 treatment. In contrast, others have shown that patients whose tumors do not express PD-L1 are good

responders.[7-9] To explain the controversies that affect PD-L1 expression, some studies have considered that the heterogeneity between axis expression and response to PD-1/PD-L1 treatment in NSCLC depends on other factors, such as more precise methods to investigate immune evasion mechanisms and the immune microenvironment, as well as greater knowledge on the immune checkpoint genomic profile and the genetic variants of the *PD-L1* gene. [10-13]

Some genetic variants have been shown to affect normal gene activation and transcriptional initiation, and hence influence the amount of mRNA and encoded protein in the cell.[14] Non-coding variants also presumably affect genetic regulatory elements, since a majority of driver variants in cancer genomes occur in non-coding regions.[15,16]

Thus, several studies have investigated the association between *PD-1* and *PD-L1* genetic variants and the risk of various cancers, but their findings have yet failed to completely elucidate this question.[17] A previous study suggested that *PD-L1* polymorphism may predict chemotherapy response and survival rates in advanced-stage NSCLC patients after first-line paclitaxel-cisplatin.[18] More recently, *PD-L1* copy number variations, point mutations, and 3′-UTR disruptions have been highlighted as genetic mechanisms of PD-L1 deregulation.[19] Furthermore, previous research on Brazilian patients suggested that their ethnic background could account for their distinct cancer's molecular profile, perhaps due to their characteristic genetic admixture, inherited from European, African, and Native American ancestors. [20-22]

We hypothesize that *PD-L1* non-coding genetic variants modulate the function of this immune checkpoint in NSCLC. To explore this issue, we investigated fifteen *PD-L1* non-coding genetic variants using next-generation sequencing (NGS) in a Brazilian cohort, aiming to uncover the effect of their ethnic admixture on NSCLC. We also combined our analyses with an *in-silico* approach to predict the impact of these genetic variants on the disease. We evaluated the associations between PD-L1 protein expression level and clinicopathological characteristics, including the prognosis of NSCLC patients undergoing surgical resection,

glimpsing the impact of genetic variants on post progression survival (PPS) and overall survival (OS). Herein, in this context, we intend to expand the existing literature on *PD-L1* gene alterations at the genetic level and their impact on NSCLC in patients from different ethnicities, thus increasing the knowledge about the molecular basis of immunotherapy biomarkers.

Material and methods

Cohort

In this retrospective multi-center study, we obtained archival formalin-fixed paraffin-embedded histologic tumor sections from 70 patients diagnosed with NSCLC (33 adenocarcinomas [ADC], 24 squamous cell carcinoma [SqCC] and 13 large cell carcinoma [LCC]) who underwent surgical resection between January 1, 1995, and December 31, 2015. Patients had been treated at the Hospital das Clínicas of the University of São Paulo Medical School (HC-FMUSP), at the Heart Institute of the University of São Paulo (INCOR), at the Cancer Institute of São Paulo (ICESP), and at the A.C. Camargo Cancer Center in São Paulo, Brazil.

All samples were histologically reviewed by lung pathologists who selected samples with at least 30% of lung cancer cells before nucleic acid extraction. The samples were classified using the 2017 International Association for the Study of Lung Cancer (IASLC) classification system.[23] The clinicopathological features of patients were obtained from the medical records. The study was approved in accordance with the ethical standards of the responsible committee on human experimentation local (Research Ethics Committee of University of São Paulo Medical School - CAAE: 79769017.1.0000.5440; opinion number: 2.673.320) and with the 1964 Helsinki declaration. A waiver of the requirement for informed consent was obtained from

committee, and to identity of the subjects under this retrospective analysis was omitted and anonymized.

Multiplex Immunofluorescence Staining

We performed a Multiplex immunofluorescence (mIF) staining using methods that had been previously described and validated.[24,25] Four-micrometer-thick consecutive TMA sections were stained using an automated staining system (BOND-RX; Leica Biosystems, Buffalo Grove, IL) to characterize PD-L1 (clone E1L3N, dilution 1:100; Cell Signaling Technology, Danvers, MA). The PD-L1 marker was stained with its respective fluorophore from the Opal 7 kit (catalogue #NEL797001KT; Akoya Biosciences/PerkinElmer, Waltham, MA). A complete validation using immunofluorescence (IF) allowed us to obtain a uniform, specific, and appropriate signal across all the channel; ie, a well-balanced staining pattern for the multiplex staining.[24,25] We also defined and optimized the correct fluorophore signals between 10 to 30 counts of intensity to maintain good balance and similar thresholds of intensity across all antibodies. In parallel, to detect possible variations in staining and optimize the separation of the signal, positive and negative (autofluorescence) controls were included during the staining process to ensure that all the antibodies performed well together. Autofluorescence controls with an expected spectral resolution of 488nm were able to accurately remove the autofluorescence from all the label signals during the analysis. The stained slides were then scanned using a multispectral microscope, the Vectra Polaris 3.0 imaging system (Akoya Biosciences/PerkinElmer, Waltham, MA), under fluorescence conditions.

Multiplex Immunofluorescence Quantitation

Multispectral images of tumor sections from each core were analyzed with inForm 2.2.1 (Akoya Biosciences/PerkinElmer, Waltham, MA) software Individual cells, which were defined by nuclei staining and identified by the InForm cell segmentation tool, were subjected

to a phenotyping pattern-recognition learning algorithm to characterize co-localization of the various cell populations using panel labeling.[26] The panel labeling was as follows: Malignant cells, (MCs) with the AE1/AE3+ marker, including those with and without PD-L1 expression (AE1/AE3+ PD-L1+ and AE1/AE3+ PD-L1-, respectively). The individual cell phenotype report produced by the InForm software was processed using Excel 2010 (Microsoft. Houston, TX), and a final summary of the data, which contained the median of each individual phenotype (given as number of cells/mm^2) and the percentage of macrophages and MCs expressing PD-L1, was created for statistical analysis. If the percentage of MCs or macrophages expressing PD-L1was greater than the median value, the PD-L1 expression was considered positive. If the percentage of macrophages or MCs expressing PD-L1 was lower than or equal to the median, the PD-L1 expression was considered negative.

DNA Extraction

Genomic DNA (gDNA) was extracted from frozen NSCLC tissue using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations. DNA concentration was measured using the Qubit® 3.0 Fluorometer (Invitrogen, Life Technologies, CA, USA). DNA integrity was assessed using the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

Sequencing and Data Analysis

We performed a *PD-L1*(*CD274*) full gene screening by deep targeted sequencing using the TruSeq Custom Amplicon Panel v1.5 kit (TSCAP, Illumina, SanDiego, CA) and the MiSeq platform (Illumina, SanDiego, CA). The DNA libraries were performed according to the manufacturer's instructions and consisted of 150 bp paired-end reads (300 cycles).

We performed an NGS data analysis on the Molecular Genetics and Bioinformatics Laboratory of the Experimental Research Unit (UNIPEX) at the Medical School of São Paulo

State University (UNESP). Sequencing quality was assessed by FastQC. Reads were aligned to the human genome (hg19, GRCh37) with BWA software, and SAM tools converted the alignment results to BAM format.[27] Next, the mapped reads underwent variant calling for SNP with GATK command line tools, including HaplotypeCaller, SelectVariants, and VariantFiltration programs with default parameters. After the calling step, the variants were annotated using the VEP[28] software. Coverage depth was a priori set at 100×. Variants had to have >10 reads of position depth (PD) and/or >6 reads of allele depth (AD) and/or an AD/PD ratio of >0.05 and/or a population frequency higher than 1% (popfreq_all >0.01) were included in the study. Finally, variants were compared using ABraOM, a web-based public database of Brazilian genomic variants.[29]

In silico Prediction of Non-coding Genetic Variants

Several tools were used to predict potential functional effects of SNPs on non-coding binding sites, such as splice sites and binding sites for transcription factors, exonic splicing enhancers (ESEs), and microRNA (miRNA). The impact of each genetic variant was assessed using VarSome[30] an integrated search engine that allows access to several databases, forecasting tools, and publications on a single website. Variant pathogenicity was reported using an automatic variant classifier that evaluates each submitted variant according to guideline of the American College of Medical Genetics and Genomics (ACMG) and classifies it as either "pathogenic", "likely pathogenic", "likely benign", "benign" or "uncertain significance". Varsome predicts the pathogenicity of each variant through a DANN[31] score, a methodology for scoring deleterious annotations of genetic variants using neural networks that results in a number ranging from 0 to 1. Higher DANN scores represent greater variant deleteriousness.[31]

Next, we applied the Genomic Evolutionary Rate Profiling (GERP)[32,33] conservation score. This score is used to calculate the reduction of substitutions in a multi-species sequence

alignment when compared to a neutral expectation. GERP scores >5.5 are strongly associated with a purifying selection. Mutations that occur at highly conserved sites in many species are assumed as harmful and therefore contribute to the genetic load within a species.

Finally, we used two tools, SNPinfo (FuncPred)[34] and RegulomeDB[35], to track SNPs according to their functions. SNPinfo is a web server that helps researchers investigate SNPs in studies of genetic association and provide different pipelines for SNP selection, whereas RegulomeDB is an online composite database and prediction tool to annotate and prioritize potential regulatory variants from the human genome.[35] RegulomeDB divides the variants into six categories: category 1 variants are likely to affect binding and are linked to the expression of a gene target, category 2 variants are likely to affect binding, category 3 variants are less likely to affect binding, and category 4, 5, and 6 variants have minimal binding evidence.[35]

Statistical Analysis

The allelic and genotypic frequencies of the *PD-L1* polymorphisms found in NSCLC were calculated by Hardy Weinberg equilibrium ([1 _ (hC 2H)]/2N, where “h" stands for a heterozygous genotype, “H" for homozygous genotype and “N" for the number of samples). Associations between polymorphisms, PD-L1 protein expression, and the clinicopathological parameters of NSCLC patients were investigated by Chi-square test. The prognostic value of each polymorphism was assessed by a survival analysis using the Kaplan–Meier method with the log-rank test for statistical significance. In addition, Cox's proportional hazards regression models were used in a multivariate analysis to test the association between SNPs and PPS and OS. PPS was considered as the period from tumor progression until death or last follow-up. OS was defined as the time from curative surgery to death or last date known to be alive. The

statistical software program IBM SPSS (version 22; Armonk, NY, USA) performed all analyses. Differences were considered statistically significant at $P<0.05$.

Results

Clinicopathologic characteristics

Of the 70 patients included in the study, 33 presented with ADC (47.1%), 24 with SqCC (34.3%), and 13 with LCC (18.6%). The clinical characteristics by histologic types are summarized in **Table 1**. While SqCC cases were more frequent in males (81.8%), ADC cases were equally distributed between genders, and LCC cases were close to equal distribution (male 55.6%, female 44.4%). All histological types were more frequent in patients aged 63 years or younger. 8 patients reported a history of tobacco smoking in the ADC group (72.7%) and in the SqCC group (88.9%), versus 3 patients in the LCC group (50.0%). All the histological subtypes included advanced stages of disease (9 cases in ADC, 6 cases in SqCC, and 4 cases in LCC). Most of the patients had not received either chemotherapy (12 cases to ADC, 8 cases to SqCC, and 6 cases to LCC) or radiation therapy (18 cases to ADC, 10 cases to SqCC, and LCC) as adjuvant treatment. Malignant cells expressed PD-L1 above the median in 7 LCC cases (70.0%), the most relevant expression compared to the other two histological subtypes. The median follow-up was 66 (12–144) months. None of the analysis revealed significant differences between histological types ($P>0.05$).

Table 1. Demographic and clinicopathological characteristics of 70 NSCLC patients.

Characteristics	Histological Subtypes Adenocarcinoma (N=33)	Squamous cell carcinoma (N=24)	Large cell carcinoma (N=13)	*P-value*
[a]Gender				
Male	10 (50.0%)	9 (81.8%)	5 (55.6%)	0.21
Female	10 (50.0%)	2 (18.2%)	4 (44.4%)	
[a]Age (years)				
≤ 63	11 (61.1%)	6 (54.5%)	6 (60.0%)	0.93
> 63	7 (38.9%)	5 (45.5%)	4 (40.0%)	
[a]Smoke Status				
Yes	8 (72.7%)	8 (88.9%)	3 (50.0%)	0.25
No	3 (27.3%)	1 (11.1%)	3 (50.0%)	
[a]Clinical Stage†				
I-II	8 (47.1%)	4 (40.0%)	3 (42.9%)	0.93
III-IV	9 (52.9%)	6 (60.0%)	4 (57.1%)	
[a]Treatment				
Chemotherapy				
Yes	7 (36.8%)	3 (27.3%)	4 (40.0%)	0.80
No	12 (63.2%)	8 (72.7%)	6 (60.0%)	
Radiotherapy				
Yes	1 (5.3%)	1 (9.1%)	0 (0.0%)	0.63
No	18 (94.7%)	10 (90.9%)	10 (100%)	
[a]Relapse				
Yes	9 (64.3%)	4 (50.0%)	2 (33.3%)	0.43
No	5 (35.7%)	4 (50.0%)	4 (66.7%)	
[a]Mcs+PD-L1 (median cell density)				
≤1.26	13 (54.2%)	10 (52.6%)	3 (30.0%)	0.40
>1.26	11 (45.8%)	9 (47.4%)	7 (70.0%)	
Follow-up (months) 66 (12-144)				
[a]Patients censored for survival analysis at last follow-up time	8 (53.3%)	4 (66.7%)	3 (42.9%)	-

Allele and genotype distributions of PD-L1 gene polymorphisms

All NSCLC patients who underwent surgical resection were successfully genotyped for fifteen *PD-L1* SNPs: rs76805387T>C, rs4742098A>G, rs47946526A>G, rs10217310G>T, rs7864231G>A, rs41280725C>T, rs573692330A>G, rs1011769981G>A, rs41280723T>C, rs138135676T>C, rs4143815G>C, rs2297136G>A, rs148242519G>A, rs41303227C>T, and rs7041009G>A. **Supplementary Table 1** shows the SNP identification numbers, allele and genotype frequencies, and *P*-value for HWE. Of the 15 SNPs studied, 11 were found to be monomorphic, whereas 4 SNPs, namely rs4742098, rs4143815, rs2297136, and rs7041009,

were polymorphic in NSCLC. Monomorphic SNPs were excluded from further analysis.

All the polymorphic SNPs were found to be in equilibrium (P>0.05) for HWE. The allele frequency of our cohort was compared to different populations in the 1000 Genomes Project (**Supplementary Table 2**).

Correlation between PD-L1 gene polymorphisms and clinicopathological characteristics in NSCLC

We performed stratified analyses on the associations between clinical characteristics and the four *PD-L1* polymorphisms with different genotypic distributions. **Table 2** shows each SNP genotype frequency and their associated clinicopathological characteristics. Three of the four *PD-L1* gene polymorphisms (rs4742098, rs4143815, and rs7041009) were significantly associated with relapse (P=0.01; P=0.05; P=0.02, respectively). For the rs4742098 variant, carriers of the G allele (AG or GG genotypes) were less likely to relapse (P=0.01). Similarly, for rs4143815, carriers of the alternative C allele (CG or CC genotypes) were also less likely to relapse (P=0.05). In rs7041009, however, carriers of the alternative allele A (AG or GG genotypes) were more likely to relapse (P=0.02). Moreover, GG genotype (reference) of rs7041009 showed a significant correlation with age, being more prevalent among younger patients (16 patients, or 69.6%), and status, being more prevalent among patients who were alive (11 patients, or 84.6%), compared to carriers of the A allele (P=0.02 and P<0.01, respectively). No statistical significance was observed in the association between rs2297136 genotypes and clinicopathological variants.

Correlation between PD-L1 gene polymorphisms and PD-L1 protein expression

The correlation between PD-L1 protein expression and *PD-L1* gene polymorphisms are shown in **Table 2**. There were no statistically significant associations between PD-L1 protein

expression in malignant cells and *PD-L1* gene polymorphisms. In our cohort, the four *PD-L1* gene polymorphisms were in non-coding regions and, apparently, cause no interference in PD-L1 protein expression in NSCLC malignant cells. However, when we correlated PD-L1 protein expression with histological subtype, we observed that the expression in malignant cells was above the median in 70% of patients with LLC, in contrast with 45.8% and 47.4% of patients with ADC and SqCC, respectively.

Table 2. Clinicopathological characteristics of 70 NSCLC patients stratified by the *PD-L1* polymorphisms rs4742098, rs4143815, rs2297136 and rs7041009

Characteristics	rs4742098 AA vs AG/GG No. of patients (N=70) (%)			rs4143815 GG vs CG/CC No. of patients (N=70) (%)			rs2297136 GG vs AG/AA No. of patients (N=70) (%)			rs7041009 GG vs AG/AA No. of patients (N=70) (%)		
	A/A	A/G+G/G	*P*	G/G	C/G+C/C	*P*	G/G	AG+A/A	*P*	G/G	A/G+A/A	*P*
Age (years), median												
≤ 63	13 (56.5%)	10 (62.5%)	0.75	11 (57.9%)	12 (60.0%)	1.00	4 (57.1%)	19 (59.4%)	1.00	16 (76.2%)	7 (38.9%)	**0.02**
>63	10 (43.5%)	6 (37.5%)		8 (42.1%)	8 (40.0%)		3 (42.9%)	13 (40.6%)		5 (23.8%)	11 (61.1%)	
Gender, n (%)												
Male	13 (56.5%)	11 (64.7%)	0.74	12 (63.2%)	12 (57.1%)	0.75	4 (57.1%)	20 (60.6%)	1.00	12 (57.1%)	12 (63.2%)	0.75
Female	10 (43.5%)	6 (35.3%)		7 (36.8%)	9 (42.9%)		3 (42.9%)	13 (39.4%)		9 (42.9%)	7 (36.8%)	
Smoke status												
Yes	10 (71.4%)	9 (75.0%)	1.00	9 (75.0%)	10 (71.4%)	1.00	4 (66.7%)	15 (75.0%)	1.00	9 (60.0%)	10 (90.9%)	0.17
No	4 (28.6%)	3 (25.0%)		3 (25.0%)	4 (28.6%)		2 (33.3%)	5 (25.0%)		6 (40.0%)	1 (9.1%)	
Histology												
ADC	18 (43.9%)	15 (51.7%)		15 (41.7%)	18 (52.9%)		7 (53.8%)	26 (45.6%)		15 (41.7%)	18 (55.9%)	
SqCC	16 (39.0%)	8 (27.6%)	0.61	15 (41.7%)	9 (26.5%)	0.40	4 (30.8%)	20 (35.1%)	0.86	12 (33.3%)	12 (35.3%)	0.34
LCC	7 (17.1%)	6 (20.7%)		6 (16.7%)	7 (20.6%)		2 (15.4%)	11 (16.3%)		9 (25.0%)	4 (11.8%)	
Clinical Stage†												
I-II	8 (40.0%)	7 (50.0%)	0.72	7 (41.2%)	8 (47.1%)	1.00	1 (14.3%)	14 (51.9%)	0.10	8 (44.4%)	7 (43.8%)	1.00
III-IV	12 (60.0%)	7 (50.0%)		10 (58.8%)	9 (52.9%)		6 (85.7%)	13 (48.1%)		10 (55.6%)	9 (56.3%)	
Mcs+PD-L1 (median)												
≤1.26	15 (57.7%)	11 (42.3%)	0.78	14 (53.8%)	12 (46.2%)	1.00	6 (23.1%)	20 (76.9%)	0.74	11 (42.3%)	15 (57.7%)	0.27
>1.26	14 (51.9%)	13 (48.1%)		14 (51.9%)	13 (48.1%)		5 (18.5%)	22 (81.5%)		16 (59.3%)	11 (40.7%)	
Chemotherapy												
Yes	7 (30.4%)	7 (41.2%)	0.52	5 (26.3%)	9 (42.9%)	0.33	3 (42.9%)	11 (33.3%)	0.67	10 (47.6%)	4 (21.1%)	0.10
No	16 (69.6%)	10 (58.8%)		14 (73.7%)	12 (57.1%)		4 (57.1%)	22 (66.7%)		11 (52.4%)	15 (78.9%)	
Radiotherapy												
Yes	1 (4.3%)	1 (5.9%)	1.00	1 (5.3%)	1 (4.8%)	1.00	1 (14.3%)	1 (3.00%)	0.32	0 (0.0%)	2 (10.5%)	0.21
No	22 (95.7%)	16 (94.1%)		18 (94.7%)	20 (95.2%)		6 (85.7%)	32 (97.0%)		21 (100.0%)	17 (89.5%)	
Relapse												
Yes	11 (78.6%)	4 (28.6%)	**0.01**	9 (75.0%)	6 (37.5%)	**0.05**	5 (71.4%)	10 (47.6%)	0.39	5 (33.3%)	10 (76.9%)	**0.02**
No	3 (21.4%)	10 (71.4%)		3 (25.0%)	10 (62.5%)		2 (28.6%)	11 (52.4%)		10 (66.7%)	3 (23.1%)	
Status												

Live	5 (38.5%)	8 (53.3%)	0.47	5 (45.5%)	8 (47.1%)	1.00	3 (50.0%)	10 (45.5%)	1.00	11 (78.6%)	2 (14.3%)	**0.00**
Dead	8 (61.5%)	7 (46.7%)		6 (54.5%)	9 (52.9%)		3 (50.0%)	12 (54.5%)		3 (21.4%)	12 (85.7%)	

Notes:

Abbreviations: ADC, Adenocarcinoma; SqCC, Squamous cell carcinoma; LCC, Large cell carcinoma; Mcs, malignant cells; PD-L1, Programmed death-ligand1

Bolded values are statically significant.

† 8th International Association for the Study of Lung Cancer [23].

Associations between PD-L1 gene polymorphisms and survival outcomes

Our first statistical test examined the individual effect of patients' characteristics to estimate statistical differences in survival using the Kaplan-Meier method (**Table 3**). Patients younger than 63 years showed increased OS, 111.62 vs 66.54 months in older patients (P=0.05). Choice of treatment was also an independent factor in diagnosis, with patients who did not receive radiotherapy presenting a better survival rate when compared with those who were treated with radiotherapy, 94.43 vs 12.00 months, respectively (P<0.01). Patients who presented disease recurrence had lower survival rates and poorer prognostic when compared with those who did not relapse, 48.23 vs 123.10 months, respectively (P<0.01).

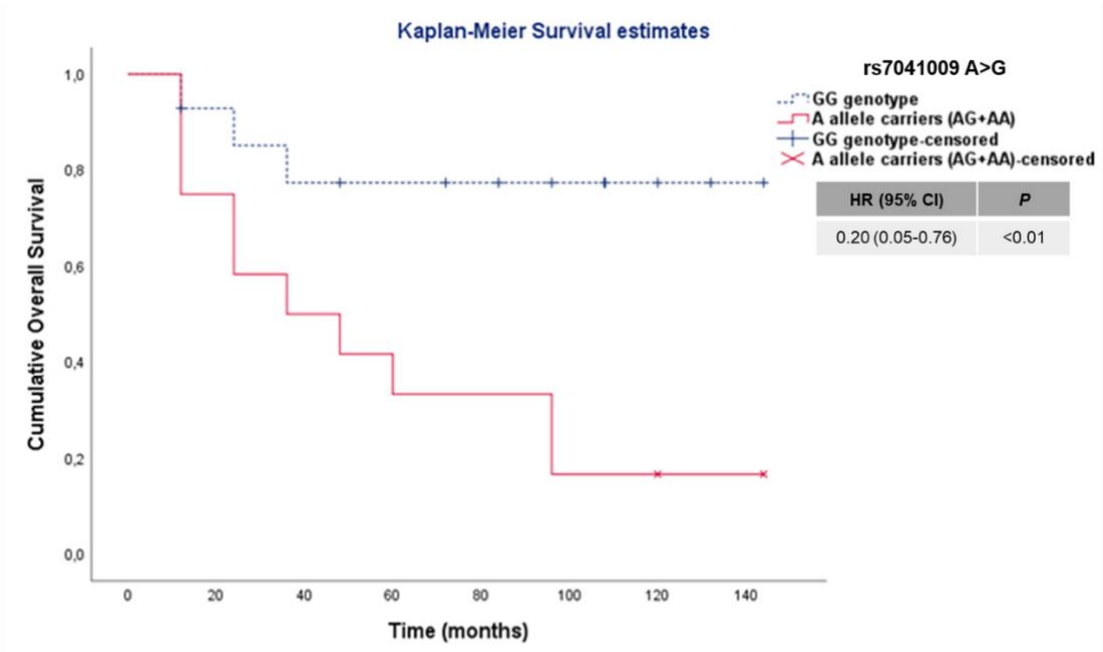
Table 3. A survival analysis conducted by the Kaplan-Meier method showing the difference in the means of the log-rank test according to the optimal upper and lower binary cut-off limits of different variables

Variables	Overall Survival (months)		Chi-Square (Log Rank)	*P*-value
	Mean	Standard Error		
Age				
≤63	111.62	13.92	3.88	0.05
>63	66.54	15.99		
Treatment Radiotherapy				
Yes	12.00	0.00	11.46	**<0.01**
No	94.43	11.31		
Relapse				
Yes	48.23	9.84	11.01	**<0.01**
No	123.10	13.48		
***PD-L1* polymorphisms rs7041009 (GG vs AG/AA genotype)**				
GG	116.93	13.81	7.49	**<0.01**
AG/AA	59.00	13.67		

Notes:
Abbreviations: PD-L1, Programmed death-ligand 1
Bolded values are statically significant.

Moreover, differences in the genotypes of *PD-L1* polymorphisms seemed to also impact the prognosis of NSCLC patients. *PD-L1* rs7041009, for instance, led to a statistically significant

difference in OS (**Figure 1)**, with carriers of the A allele of rs7041009 having lower OS than carriers of the GG genotype (reference), 59.00 vs 116.93 months, respectively (*P*<0.01).



rs7041009 G>A	Survival estimates (months)											
	12	24	36	48	60	72	84	96	108	120	132	144
GG genotype	60.0%	33.3%	50.0%	50.0%	0.0%	100.0%	100.0%	33.3%	100.0%	50.0%	100.0%	50.0%
AG genotype	40.0%	33.3%	50.0%	50.0%	100.0%	0.0%	0.0%	67.7%	0.0%	50.0%	0.0%	0.0%
AA genotype	0.0%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	50.0%

Figure 1. Kaplan-Meier survival curve for *PD-L1* rs7041009 G>A. A allele carriers (AG+AA) presented worse prognosis and a lower survival rate when compared to GG genotyped patients (*P*<0.01).

Next, using a univariate Cox Regression analysis, we were able to associate the following variables with a lower risk of death: the absence of radiotherapy treatment, relapse, and GG genotype of *PDL1* rs7041009 (**Table 4**). However, after feeding these variables into a multivariate analysis, only the absence of radiotherapy treatment and relapse were considered to be independent factors for OS (HR 9.82, *P*=0.02; HR 6.15, *P*=0.04, respectively).

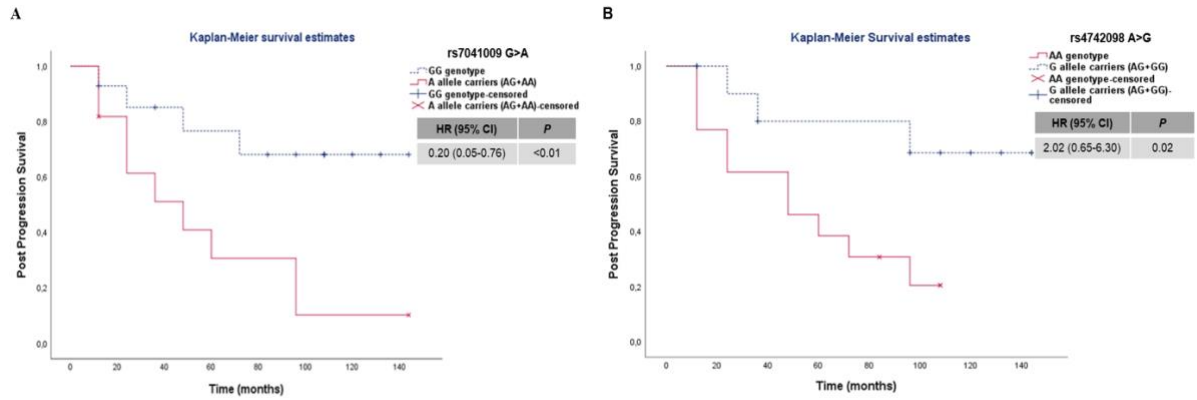
Table 4. Variables associated with overall survival (OS) in 70 patients diagnosed with NSCLC. Univariate and multivariate analyses employing a Cox proportional hazards model.

Clinicopathological Characteristics	Univariate Analysis[a]			Multivariate Analysis[b]	
	HR[c] (95% CI)[d]	HR[c]	*P* value	HR[c] (95% CI)[d]	*P*-value
Age (yrs): ≤63 vs >63	0.33 (0.10-1.10)	-1.10	0.07		
Gender					
Male	1.83 (0.54-6.14)	0.60	0.32		
Smoke status					
Yes	1.33 (0.33-5.36)	0.28	0.68		
Histology					
ADC	1.25 (0.32-4.86)	0.22	0.74		
SqCC	1.67 (0.33-8.37)	0.51	0.53		
LCC (reference)	-	-	0.82		
Clinical Stage†					
I-II	0.41 (0.11-1.52)	-0.87	0.18		
III-IV (reference)	-	-	-		
Therapy					
Chemotherapy Yes	0.36 (0.10-1.34)	-1.00	0.13		
Relapse *Yes*	8.47 (1.81-39.62)	2.13	**0.00**	6.15 (1.03-36.44)	**0.04**
Immune checkpoint					
PD-L1+/mm²	1.00 (0.95-1.05)	0.00	0.94		
polymorphisms in *PD-L1*					
rs4742098					
AA genotype	2.02 (0.65-6.30)	0.70	0.22		
G allele carrier (reference)	-	-	-		
rs4143815					
GG genotype	1.58 (0.52-4.81)	0.46	0.41		
C allele carrier (reference)	-	-	-		
rs2297136					
GG genotype	1.24 (0.34-4.55)	0.22	0.73		
A allele carrier (reference)	-	-	-		
rs7041009					
GG genotype	0.20 (0.05-0.76)	-1.56	**0.01**	0.57 (0.12-2.70)	0.48
A allele carrier (reference)	-	-	-		

Notes:
Abbreviations: ADC, Adenocarcinoma; SqCC, Squamous cell carcinoma; LCC, Large cell carcinoma; HR= hazard ratio (β coefficient); CI=confidence interval; PD-L1, Programmed death-ligand 1
Bolded values are statically significant.
† 8th International Association for the Study of Lung Cancer [23].

Then, we introduced the *PD-L1* polymorphisms into the Cox model, controlling for radiotherapy treatment and tumor relapse. Of the four SNPs, only rs7041009 was identified as a co-dependent factor associated with radiotherapy and relapse. We thus inferred that patients with

NSCLC who carried the A allele (AG/AA) presented a higher risk of relapse in the presence of radiotherapy, resulting in a poorer prognosis and decreased survival rates than patients who carried the rs7041009 GG genotype. In relapsed patients, we observed that the *PD-L1* polymorphisms rs7041009 and rs4742098 might have an impact on PPS (**Figure 2**). Patients with the rs7041009 GG genotype had a higher PPS than those with the alternative A allele of rs7041009 (AG/AA), 110.98 vs 56.18 months, respectively (P<0.01); whereas, patients who carried the reference rs4742098 AA genotype had lower PPS than those who carried the alternative G allele of rs4742098 (AG/GG), 56.00 vs 115.71 months, respectively (P=0.02).



rs7041009 G>A	Post Progression Survival estimates (months)						
	12	24	36	48	60	72	96
GG genotype	50.0%	33.3%	0.0%	50.0%	0.0%	100.0%	0.0%
AG genotype	50.0%	33.3%	100.0%	50.0%	100.0%	0.0%	100.0%
AA genotype	0.0%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%

rs4742098 A>G	Post Progression Survival estimates (months)						
	12	24	36	48	60	72	96
AA genotype	75.0%	66.7%	0.0%	100.0%	100.0%	100.0%	50.0%
AG genotype	25.0%	33.3%	0.0%	0.0%	0.0%	0.0%	50.0%
GG genotype	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%

Figure 2. Kaplan-Meier survival curves estimating post-progression survival (PPS) in NSCLC patients according to *PD-L1* polymorphisms. **A)** Kaplan-Meier survival curve for rs7041009 G>A. A allele carriers (AG+AA) presented worse prognosis and a lower PPS rate when compared GG genotyped patients (P<0.01); **B)** Kaplan-Meier survival curve for rs4742098 A>G. G allele carriers (AG+GG) had a higher PPS rate and better prognosis when compared AA genotyped patients (P=0.02).

In silico prediction of PD-L1 gene polymorphisms

The *in silico* analysis predicted the *PD-L1* variants rs4742098 (c.*2635A>G), rs4143815 (c.*395G>C), rs2297136 (c.*93G>A), and rs7041009(c.682+122G>A) to be benign (**Table 5**). Not only was the DANN score low for all four variants (0.8226, 0.6475, 0.7056, and 0.5428, respectively), but their GERP score was also lower than 5.5 (-1.74, 2.38, 4.4, and 1.81,

respectively), indicating that these variants are found in non-conserved positions and are unlikely to be harmful.

Table 5. List of the selected nonconding SNP and the tools used to study them.

dbSNP ID	Region	FuncPred	RegulomeDB score	VarSome (DANN Score)
rs4742098	3'UTR	--	5	Benign (0.8226)
rs4143815	3'UTR	miRNA-binding function	5	Benign (0.6475)
rs2297136	3'UTR	miRNA-binding function	5	Benign (0.7056)
rs7041009	Intron	--	6	Benign (0.5428)

Notes:
Abbreviations: SNPs: Single‑nucleotide polymorphisms; dbSNP: SNP database; FuncPred: SNP Function Prediction; RegulomeDB score: RegulomeDB variant classification; VarSome: The Human Genomics Community; DANN Score: Annotation of genetic variants using neural networks.
-- SNPs that don't affect function.

SNPinfo predicted miRNA-binding function to be affected by two of these variants, namely rs2297136 and rs4143815. Rs2297136 was predicted to affect the binding function of hsa-miR-324-5p and hsa-miR-632, whereas rs4143815 was found to correlate with hsa-miR-1252, hsa-miR-1253, hsa-miR-539, hsa-miR-548, and hsa-miR-570 (**Table 6**). RegulomeDB was then used to complement the SNP analysis. Three of the four SNPs, rs4742098, rs4143815, and rs2297136, had a RegulomeDB score of 5, whereas rs7041009 had a score of 6, meaning that all four variants show minimal binding evidence (**Table 5**).

Table 6. List of the 3′ UTR SNPs analyzed in FuncPred and their miRNA motif

dbSNP ID	Allele	miRNA motif
rs4143815	C	*hsa-miR-1252*
	C	*hsa-miR-1253*
	C	*hsa-miR-539*
	C	*hsa-miR-548*
	G	*hsa-miR-570*
rs2297136	A/G	*hsa-miR-324-5p*
	A/G	*hsa-miR-632*

Notes:
Abbreviations: FuncPred: SNP Function Prediction; dbSNP ID: SNP database identification.

Discussion

Lung cancer has a high mortality rate and lacks suitable markers for early diagnosis and prognosis. Thus, it is essential to detect the best potential biomarker out of the several genetic and protein markers. Fortunately for patients, the translational impact of such findings is rapidly increasing, and the stimulation of immune response by ICIs has emerged as a dramatic paradigm shift in the treatment of advanced tumors, mainly NSCLC.[19] [36] PD-1/PD-L1 monoclonal antibodies have shown potential efficacy in advanced squamous-cell and non-squamous NSCLC.[37] [38] However, despite the remarkable success achieved by immunotherapy so far, its effectiveness still seems to vary among cancer patients.[19] The expression of PD-L1 on tumor cells remains the only recognized predictive factor for immunotherapy response in NSCLC patients; however patients without PD-L1 expression on tumor cells may also respond to immunotherapy.[39-43] Based on these findings, the present study inferred that *PD-L1* non-coding genetic variants could help predict the prognosis of patients with NSCLC and impact disease recurrence and OS.

In our cohort of 70 NSCLC specimens, we evaluated PD-L1 protein expression in malignant cells by PD-L1 multiplex immunofluorescence (mIF) assays, using the Cell Signaling E1L3N clone and image analysis, and investigated *PD-L1* polymorphisms by NGS sequencing. This method resulted in the detection of high PD-L1 expression in LCC malignant cells when compared to other histological subtypes, suggesting that LCC patients may benefit from ICIs. As described by Shimoji et al.,[44] PD-L1 expression using the Cell Signaling E1L3N clone was significantly correlated with a consistent vimentin expression, increased Ki-67 labeling index and poor prognosis in ADC but not in SqCC. Other studies have reported that PD-L1 was detected at significantly higher frequencies in SqCC than in ADC of the lung.[44] [45] Cha et al.[46] found that PD-L1 expression using the SP142 clone was significantly associated with an ADC solid subtype histology, p53 aberrant expression, and poor prognosis.

We also assessed *PD-L1* polymorphisms. Of the fifteen genetic variants genotyped, eleven were monomorphic. The four potential variants (rs4742098, rs4143815, rs2297136, and rs7041009) present in the non-coding region were correlated with the clinicopathological characteristics of the NSCLC patients and were submitted to an *in silico* analysis investigating their functional role. The MAF of the rs4742098, rs4143815, and rs7041009 polymorphisms was consistent with the genotype frequency among European and Mixed Americans populations present in The 1000 Genomes Project, whereas the G allele in the rs2297136 polymorphism was the main allele in our cohort to show racial differences.

When assessing patient prognosis, three of the four *PD-L1* variants rs4742098, rs4143815, and rs7041009 were significantly associated with disease recurrence. Carriers of the G allele (individuals with the AG or GG genotypes) of rs4742098 were less likely to relapse compared to carriers of the homozygous AA genotype. Similar findings were published by Du and colleagues,[47] who reported that the AG genotype differed from the AA genotype in terms of risk of NSCLC recurrence. In the case of rs4143815G, patients with the alternative C allele were less likely to relapse in our study, in agreement with Nomizo et al.'s report[48]. In their study, the authors even suggested that this polymorphism might be a biomarker for nivolumab efficacy.[48] Finally, in our study, for rs7041009G>A, carriers of the alternative G allele were more likely to exhibit relapse. The rs7041009 GG genotype also showed a significant correlation with age, being more present in younger patients, and with status, being more present in patients who are alive, compared to carriers of the A allele. Rs7041009 (c.682+122G>A) is located at position 2377 in intron 4 of the PD-L1 gene. However, little is known about the exact function of this genetic variation, except that it is located near the transcription factor binding site.

Our cohort showed a significant association between these PD-L1 polymorphisms and OS in NSCLC. 18 Patients presenting the GG genotype of rs7041009 benefited from longer OS. In addition, our findings indicated that both the rs7041009G and rs4742098G polymorphisms were

significantly related to a longer PPS. The clinical impacts of *PD-L1* variants had also been investigated by previous studies. Zhao et al.[49] suggested that patients with the GG genotype of another *PD-L1* polymorphism (rs822336) had worse disease-free survival and OS in a Chinese patient population.

Further contribution was provided in that respect by Lee et al.[50], who demonstrated that rs4143815 and rs2297136 were significantly associated with clinical outcomes after chemotherapy. Of 379 patients with NSCLC treated with first-line paclitaxel–cisplatin chemotherapy, those carrying the rs4143815 G allele responded better to chemotherapy and had gains in overall survival. In our study, however, the polymorphism rs2297136 showed no significant association with clinical outcome, a finding that was corroborated by Zhao et al.[49] This difference might be explained by the heterogeneity of patients enrolled in each study, and further research is needed to settle this question.

In our study, we were unable to find any statistical significance between the rs4742098, rs4143815, rs2297136, and rs7041009 genotypes and PD-L1 protein expression in malignant cells. However, recent data have helped to shed light on the impact of *PD-L1* genetics on PD-L1 expression, though the existing results remain controversial. Recently, Tao and colleagues[51]showed that rs4143815 and rs10815225 in the *PD-L1* gene contributed to PD-L1 overexpression in gastric cancer. So far, Lee et al. [18] have conducted the largest study on *PD-L1* polymorphisms and PD-L1 expression in NSCLC. The authors showed that rs822336C, rs822337A and rs4143815G were associated with worse OS in NSCLC patients but found no significant correlation between PD-L1 expression and the genotypes of these polymorphisms. Krawczyk et al. [52] demonstrated that carriers of the rs822335 CC genotype were predisposed to higher expression of the PD-L1 protein in NSCLC tumor cells, whereas rs822336 had no effect on PD-L1 expression in these cells. Their results are consistent with our findings, but future research is needed to clarify remaining confounders.

In our study, using *in silico* approaches, we report that rs4742098, rs4143815, rs2297136, and rs7041009 can be considered benign variants. However, little is known about the effect that mutations in conserved non-coding regions might have on fitness and how many of them are present in the human genome as deleterious polymorphisms. Moreover, rs2297136 was predicted to affect the miRNA-binding function of hsa-miR-324-5p and hsa-miR-632, whereas the variant rs4143815 was found to correlate with hsa-miR-1252, hsa-miR-1253, hsa-miR-539, hsa-miR-548, and hsa-miR-570. In this context, others have reported that variants in the 3′UTR region of the *PD-L1* gene can affect the interaction of miRNAs, possibly resulting in *PD-L1* underexpression.[53] Therefore, additional studies are necessary to validate our findings. However, there are limitations to our analysis. We did not perform a case-control approach and our cohort comprises a relatively small sample size. Nonetheless, to our knowledge, our research on the effect of the rs7041009 polymorphism of *PD-L1* gene on NSCLC patients is unique.

We believe this study is also the first to evaluate variants in the non-coding region of the *PD-L1* in Brazilian patients with NSCLC, since most studies of *PD-L1* polymorphisms have been conducted in Asian patients. Thus, we consider this exploratory study as a pioneer in the understanding of *PD-L1* polymorphisms in a genetic admixed population.

References

1. Shankar A, Saini D, Dubey A, et al. Feasibility of lung cancer screening in developing countries: challenges, opportunities and way forward. *Transl Lung Cancer Res*. 2019;8(Suppl 1):S106-S121.
2. Berghmans T, Durieux V, Hendriks LEL, Dingemans AM. Immunotherapy: From Advanced NSCLC to Early Stages, an Evolving Concept. *Front Med (Lausanne)*. 2020;7:90.
3. Bylicki O, Paleiron N, Rousseau-Bussac G, Chouaïd C. New PDL1 inhibitors for non-small cell lung cancer: focus on pembrolizumab. *Onco Targets Ther*. 2018;11:4051-4064.

4. Yeo MK, Choi SY, Seong IO, Suh KS, Kim JM, Kim KH. Association of PD-L1 expression and PD-L1 gene polymorphism with poor prognosis in lung adenocarcinoma and squamous cell carcinoma. *Hum Pathol.* 2017;68:103-111.
5. Francisco LM, Salinas VH, Brown KE, et al. PD-L1 regulates the development, maintenance, and function of induced regulatory T cells. *J Exp Med.* 2009;206(13):3015-3029.
6. Lantuejoul S, Damotte D, Hofman V, Adam J. Programmed death ligand 1 immunohistochemistry in non-small cell lung carcinoma. *J Thorac Dis.* 2019;11(Suppl 1):S89-S101.
7. Aggarwal C, Abreu DR, Felip E, et al. Prevalence of PD-L1 expression in patients with non-small cell lung cancer screened for enrollment in KEYNOTE-001, -010, and -024. *Ann Oncol.* 2016;27(6):359-378.
8. Yu H, Boyle TA, Zhou C, Rimm DL, Hirsch FR. PD-L1 Expression in Lung Cancer. *J Thorac Oncol.* 2016;11(7):964-975.
9. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med.* 2016;375:1823-1833.
10. Marwitz S, Scheufele S, Perner S, Reck M, Ammerpohl O, Goldmann T. Epigenetic modifications of the immune-checkpoint genes CTLA4 and PDCD1 in non-small cell lung cancer results in increased expression. *Clin Epigenetics.* 2017;9:51.
11. Chen L, Gibbons DL, Goswami S, et al. Metastasis is regulated via microRNA-200/ZEB1 axis control of tumour cell PD-L1 expression and intratumoral immunosuppression. *Nat Commun.* 2014;5:5241.
12. Ma Y, Adjemian S, Mattarollo SR, et al. Anticancer chemotherapy-induced intratumoral recruitment and differentiation of antigen-presenting cells. *Immunity.* 2013;38(4):729-741.
13. Mazzaschi G, Madeddu D, Falco A, et al. Low PD-1 Expression in Cytotoxic CD8(+) Tumor-Infiltrating Lymphocytes Confers an Immune-Privileged Tissue Microenvironment in NSCLC with a Prognostic and Predictive Value. *Clin Cancer Res.* 2018;24(2):407-419.
14. de Vooght KMK, van Wijk R, van Solinge WW. Management of Gene Promoter Mutations in Molecular Diagnostics. *Clin Chem.* 2009;55(4):698-708.
15. Cuykendall TN, Rubin MA, Khurana E. Non-coding genetic variation in cancer. *Curr Opin Syst Biol.* 2017;1:9-15.

16. Amlie-Wolf A, Tang M, Way J, et al. Inferring the Molecular Mechanisms of Noncoding Alzheimer's Disease - Associated Genetic Variants. *J Alzheimers Dis.* 2019;72(1):301-318.
17. Hashemi M, Karami S, Sarabandi S, et al. Association between PD-1 and PD-L1 Polymorphisms and the Risk of Cancer: A Meta-Analysis of Case-Control Studies. *Cancers (Basel).* 2019;11(8):1150.
18. Lee SY, Jung DK, Choi JE, et al. Functional polymorphisms in PD-L1 gene are associated with the prognosis of patients with early stage non-small cell lung cancer. *Gene.* 2017;599:28-35.
19. Fabrizio FA, Trombetta D, Rossi A, Sparaneo AA, Castellana S, Muscarella LA. Gene code CD274/PD-L1: from molecular basis toward cancer immunotherapy. *Ther Adv Med Oncol*. 2018;10:1758835918815598.
20. Araujo LA, Baldotto CA, Castro GJA, et al. Lung cancer in Brazil. *J Bras Pneumol.* 2018; 44(1):55-64.
21. de Melo AC, de Sá VK, Sternberg C, et al. Mutational Profile and New IASLC/ATS/ERS Classification Provide Additional Prognostic Information about Lung Adenocarcinoma: A Study of 125 Patients from Brazil. *Oncology.* 2015;89(3):175-86.
22. de Sá VK, Coelho JC, Capelozzi VL, de Azevedo SJ. Lung cancer in Brazil: epidemiology and treatment challenges. *Lung Cancer.* 2016;7:141-148.
23. Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol.* 2016;11(1):39-51.
24. Parra ER, Uraoka N, Jiang M, et al. Validation of multiplex immunofluorescence panels using multispectral microscopy for immune-profiling of formalin-fixed and paraffin-embedded human tumor tissues. *Sci Rep*. 2017;7(1):13380.
25. Parra ER, Jiang M, Machado-Rugolo J, et al. Variants in Epithelial-Mesenchymal Transition and Immune Checkpoint Genes Are Associated With Immune Cell Profiles and Predict Survival in Non-Small Cell Lung Cancer. *Arch Pathol Lab Med.* 2020; 144(10):1234-1244.
26. Gorris MAJ, Halilovic A, Rabold K, et al. Eight-Color Multiplex Immunohistochemistry for Simultaneous Detection of Multiple Immune Checkpoint Molecules within the Tumor Microenvironment. *J Immunol.* 2018;200(1):347-354.
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.

28. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17:122.
29. Naslavsky MA, Yamamoto GL, de Almeida TF, et al. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. *Hum Mutat*. 2017;38(7):751-763.
30. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. *Bioinformatics.* 2019; 35(11):1978–1980.
31. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*. 2015;31(5):761-763.
32. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol*. 2010; 6(12):e1001025.
33. Cooper GM, Stone EA, Asimenos G, et al. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* 2005;15(7):901-913.
34. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res*. 2009;37(Web Server issue):W600-W605.
35. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-1797.
36. Nishino M, Ramaiya NH, Hatabu H, Hodi FS. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol.* 2017;14(11):655–668.
37. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer. *N Engl J Med.* 2015;373(17):1627-1639.
38. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non–Small-Cell Lung Cancer. *N Engl J Med.* 2015;373(2):123-135.
39. Yi C, He Y, Xia H, Zhang H, Zhang P. Review and perspective on adjuvant and neoadjuvant immunotherapies in NSCLC. *Onco Targets Ther*. 2019;12:7329-7336.
40. Horn L, Spigel DR, Vokes EE, et al. Nivolumab Versus Docetaxel in Previously Treated Patients With Advanced Non-Small-Cell Lung Cancer: Two-Year Outcomes From Two Randomized, Open-Label, Phase III Trials (CheckMate 017 and CheckMate 057). *J Clin Oncol.* 2017;35(35):3924-3933.
41. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer. *N Engl J Med.* 2017;377(20):1919-1929.

42. Fehrenbacher L, von Pawel J, Park K, et al. Updated Efficacy Analysis Including Secondary Population Results for OAK: A Randomized Phase III Study of Atezolizumab versus Docetaxel in Patients with Previously Treated Advanced Non-Small Cell Lung Cancer. *J Thorac Oncol*. 2018;13(8):1156-1170.
43. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer. *N Engl J Med.* 2018;378(22):2078-2092.
44. Shimoji M, Shimizu S, Sato K, et al. Clinical and pathologic features of lung cancer expressing programmed cell death ligand 1 (PD-L1). *Lung Cancer*. 2016;98:69-75.
45. Takada K, Toyokawa G, Okamoto T, et al. A Comprehensive Analysis of Programmed Cell Death Ligand-1 Expression With the Clone SP142 Antibody in Non-Small-Cell Lung Cancer Patients. *Clin Lung Cancer*. 2017;18(5):572-582.
46. Cha YJ, Kim HR, Lee CY, Cho BC, Shim HS. Clinicopathological and prognostic significance of programmed cell death ligand-1 expression in lung adenocarcinoma and its relationship with p53 status. *Lung Cancer*. 2016;97:73-80.
47. Du W, Zhu J, Chen Y, et al. Variant SNPs at the microRNA complementary site in the B7‐H1 3'‐untranslated region increase the risk of non‐small cell lung cancer. *Mol Med Rep*. 2017;16(3):2682-2690.
48. Nomizo T, Ozasa H, Tsuji T, et al. Clinical Impact of Single Nucleotide Polymorphism in PD-L1 on Response to Nivolumab for Advanced Non-Small-Cell Lung Cancer Patients. *Sci Rep.* 2017;7(1):45124.
49. Zhao M, Zhang J, Chen S, Wang Y, Tian Q. Influence of Programmed Death Ligand-1-Gene Polymorphism rs822336 on the Prognosis and Safety of Postoperative Patients with NSCLC Who Received Platinum-Based Adjuvant Chemotherapy. *Cancer Manag Res*. 2020;12:6755-6766.
50. Lee SY, Jung DK, Choi JE, et al. PD-L1 polymorphism can predict clinical outcomes of non-small cell lung cancer patients treated with first-line paclitaxel-cisplatin chemotherapy. *Sci Rep*. 2016;6:25952.
51. Tao LH, Zhou XR, Li FC, et al. A polymorphism in the promoter region of PD-L1 serves as a binding-site for SP1 and is associated with PD-L1 overexpression and increased occurrence of gastric cancer. *Cancer Immunol Immunother*. 2017;66(3):309–318.
52. Krawczyk P, Grenda A, Wojas-Krawczyk K, et al. PD-L1 gene copy number and promoter polymorphisms regulate PD-L1 expression in tumor cells of non-small cell lung cancer patients. *Cancer Genet*. 2019;237:10-8.

53. Yu Z, Li Z, Jolicoeur N, et al. Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. *Nucleic Acids Res.* 2007;35(13):4535-4541.

Supplementary Table 1: Allele Frequency of Genotyped Single Nucleotide Polymorphisms (SNPs) Present in the Study and in Different Populations of the 1000 Genomes Project.

Gene	Genetic variant	Chromosome position*	Alleles	Function	EUR **		AFR**		AME**		NSCLC patients	
					Ref allele	Alt allele	Ref allele	Alt allele	Ref allele	Alt allele	Ref allele	Alt allele
	rs76805387	9:5470775	T > C	Intergenic	T=0.9334	C=0.0666	T=0.9970	C=0.0030	T=0.981	C=0.019	T=1.00	C=0.00
	rs4742098	9:5470497	A > G	3'UTR	A=0.7008	G=0.2992	A=0.8865	G=0.1135	A=0.605	G=0.395	A=0.76	G=0.23
	rs147946526	9:5470356	A > G,C	3'UTR	A=1.0000	C=0.0000	A=1.0000	C=0.0000	A=0.988	C=0.012	A=1.00	C=0.00
	rs10217310	9:5470242	G > T	3'UTR	G=1.0000	T=0.0000	G=0.8495	T=0.1505	G=0.983	T=0.017	G=1.00	T=0.00
	rs7864231	9:5469669	G > A	3'UTR	G=0.9970	A=0.0030	G=0.8359	A=0.1641	G=0.980	A=0.020	G=1.00	A=0.00
	rs41280725	9:5469654	C > T	3'UTR	C=0.9851	T=0.0149	C=1.0000	T=0.0000	C=0.993	T=0.007	C=1.00	T=0.00
PD-L1	rs573692330	9:5469493	A > G	3'UTR	A=1.0000	G=0.0000	A=1.0000	G=0.0000	A=0.999	G=0.001	A=1.00	G=0.00
	rs1011769981	9:5468898	G > A	3'UTR	NA	NA	NA	NA	NA	NA	G=1.00	A=0.00
	rs41280723	9:5468647	T > C	3'UTR	T=0.9751	C=0.0249	T=0.9970	C=0.0030	T=0.984	C=0.016	T=1.00	C=0.00
	rs138135676	9:5468535	T > C	3'UTR	T=1.0000	C=0.0000	T=0.9887	C=0.0113	T=1.000	C=0.000	T=1.00	C=0.00
	rs4143815	9:5468257	G > C	3'UTR	G=0.6700	C=0.3300	G=0.9879	C=0.0121	G=0.575	C=0.425	G=0.72	C=0.28
	rs2297136	9:5467955	G > A	3'UTR	G=0.4533	A=0.5467	G=0.3419	A=0.6581	G=0.344	A=0.656	G=0.65	A=0.35
	rs148242519	9:5467925	G > A	3'UTR	G=1.0000	A=0.0000	G=0.9705	A=0.0295	G=0.996	A=0.004	G=1.00	A=0.00
	rs41303227	9:5467801	C > T	Intron	C=0.9404	T=0.0596	C=0.9879	T=0.0121	C=0.984	T=0.016	C=1.00	T=0.00
	rs7041009	9:5463243	G > A	Intron	G=0.7117	A=0.2883	G=0.4062	A=0.5938	G=0.573	A=0.427	G=0.73	A=0.27

Notes:

Abbreviations:

*chromosomal position according to GRCh37 genome assembly.

**allele frequency from 1000 Genomes Project.

EUR: European; AFR: African; AMR: Mixed Americans; NA: not available.

Supplementary Table 2. Genotype, Allele Distributions and HWE Test for SNPs in 70 Patients Diagnosed with NSCLC.

Gene	SNP Id	Genotype frequency (%)			Allele frequency		HWE
		TT	**TC**	**CC**	**T**	**C**	
	rs76805387	70 (100)	0 (0)	0 (0)	1	0	monomorphic
		AA	**AG**	**GG**	**A**	**G**	
	rs4742098	41 (58.6)	22 (31.4)	7 (10)	0.7	0.2	2.2
		AA	**AG**	**GG**	**A**	**G**	
	rs147946526	70 (100)	0 (0)	0 (0)	1	0	monomorphic
		GG	**GT**	**TT**	**A**	**G**	
	rs10217310	70 (100)	0 (0)	0 (0)	1	0	monomorphic
		GG	**GA**	**AA**	**G**	**A**	
	rs7864231	70 (100)	0 (0)	0 (0)	1	0	monomorphic
		CC	**CT**	**TT**	**C**	**T**	
PD-L1	rs41280725	70 (100)	0 (0)	0 (0)	1	0	monomorphic
		AA	**AG**	**GG**	**A**	**G**	
	rs573692330	70 (100)	0 (0)	0 (0)	1	0	monomorphic
		GG	**GA**	**AA**	**A**	**G**	
	rs1011769981	70 (100)	0 (0)	0 (0)	1	0	monomorphic
		TT	**TC**	**CC**	**T**	**C**	
	rs41280723	70 (100)	0 (0)	0 (0)	1	0	monomorphic
		TT	**TC**	**CC**	**T**	**C**	
	rs138135676	70 (100)	0 (0)	0 (0)	1	0	monomorphic
		GG	**GC**	**CC**	**G**	**C**	
	rs4143815	36 (51.4)	27 (38.6)	7 (10)	0.7	0.2	0.3
		GG	**GA**	**AA**	**G**	**A**	
	rs2297136	13 (18.6)	26 (37.1)	31 (44.3)	0.3	0.6	2.9
		GG	**GA**	**AA**	**G**	**A**	
	rs148242519	70 (100)	0 (0)	0 (0)	1	0	monomorphic
		CC	**CT**	**TT**	**C**	**T**	
	rs41303227	70 (100)	0 (0)	0 (0)	1	0	monomorphic
		GG	**GA**	**AA**	**G**	**A**	
	rs7041009	37 (52.9)	26 (37.1)	7 (10)	0.7	0.2	0.5

Abbreviations: *PD-L1*, programmed death-ligand 1; SNPs, single nucleotide polymorphisms; HWE, Hardy-Weinberg equilibrium.

5. 3 Artigo 3: citação completa (ANEXO 3)

Aceito para publicação no *Brazilian Journal of Medical and Biological Research* (ISSN 1414-431X) em outubro de 2022.

Fator de impacto = 2.9

Clinical outcome of Brazilian patients with non-small cell lung cancer in early stage harboring rare epidermal growth factor receptor mutations

Juliana Machado-Rugolo[1,2], Camila Machado Baldavira[1], Tabatha Gutierrez Prieto[1]Eloisa Helena Ribeiro Olivieri[3], Alexandre Todorovic Fabro[1,4,] Claudia Aparecida Rainho[5,6], Erick Cruz Castelli[7,8], Paulo Eduardo Martins Ribolla[8,9], Alexandre Muxfeldt Ab'Saber[1], Teresa Takagaki[10], Maria Aparecida Nagai[11,12], Vera Luiza Capelozzi[1]

[1] Laboratory of Histomorphometry and Lung Genomics, Department of Pathology, Medical School, University of São Paulo (USP), São Paulo, Brazil

[2] Health Technology Assessment Center, Clinical Hospital of Botucatu, Medical School, São Paulo State University (Unesp), Botucatu, Brazil

[3]International Center of Research/CIPE, AC Camargo Cancer Center, São Paulo, São Paulo, Brazil

[4]Department of Pathology and Legal Medicine, Respiratory Medicine Laboratory, Ribeirão Preto Medical School, University of São Paulo (USP), São Paulo, Brazil

[5] Institute of Biosciences, Department of Chemical and Biological Sciences, São Paulo State University (Unesp), Botucatu, Brazil

[6]Molecular Genetics ad Bioinformatics Laboratory, Experimental Research Unit (UNIPEX), Medical School, São Paulo State University (Unesp), Botucatu, Brazil

[7]Department of Pathology, Medical School, São Paulo State University (Unesp), Botucatu, Brazil

[8]Biotechnology Institute (IBTEC), São Paulo State University (Unesp), Botucatu, Brazil

[9]Institute of Biosciences, Department of Biostatistics, Plant Biology, Parasitology and Zoology, State University (Unesp), Botucatu, Brazil

[10]Division of Pneumology, Instituto do Coração (Incor), University of São Paulo Medical School (USP), São Paulo, Brazil

[11]Department of Radiology and Oncology, Medical School, University of São Paulo (USP), São Paulo, Brazil

[12]Laboratory of Molecular Genetics, Center for Translational Research in Oncology, Cancer Institute of São Paulo (ICESP), São Paulo, Brazil

Abstract

Introduction: The common epidermal growth factor receptor (EGFR) mutations, such as the L858R point mutation in exon 21 and the in-frame deletional mutation in exon 19, have been definitively associated with response to EGFR-tyrosine kinase inhibitors (EGFR-TKI). However, the clinical outcome and response to treatment for many other rarer mutations are still unclear. In this study, we report the results of Brazilian patients in stage IB–IIIA non-small cell lung cancer (NSCLC) following complete resection with minimal residual disease and EGFR mutations treated with adjuvant chemotherapy and/or EGFR-TKIs.

Methods: The frequency of EGFR mutations was investigated in 70 cases of NSCLC early stage. Mutations in exons 18 and 20, uncommon mutations in exons 19 and 21, as well as in exons 3, 7, 14, 16, 22, 27, and 28, and/or the presence of different mutations in a single tumor (complex mutations) was considered rare.

Results: EGFR mutations were detected in 23 tumors (32.9%). Fourteen cases carried rare mutations and were treated with platin-based chemotherapy and two with erlotinib. The clinical outcome was described case by case with references to the literature. Notably, we found two rare EGFR mutations and one of them with an unknown response to chemotherapy and/or EGFR-TKIs.

Conclusions: We have provided complementary information concerning the clinical outcome and treatment in patients with NSCLC early stage for several rare EGFR mutations not previously or only rarely reported. Description of cases harboring rare mutations can provide the decision-making procedure in this subset of patients.

Keywords: Lung cancer, next generation sequency, rare EGFR mutation, platin-based chemotherapy, erlotinib.

Introduction

In Brazil, the latest 2021 tumor registration data show that there were approximately 17.760 new cases of lung cancer in men and 12.440 in women (1). Non-small cell lung cancer (NSCLC) accounts for 85% of lung cancer in Brazilian patients [1]. Almost 20% of them have early-stage disease (stage I–II) and around 30% of patients have locally advanced disease (stage III) at the time of diagnosis of NSCLC (2). Operative resection is the gold standard curative treatment for patients with NSCLC; nevertheless, local or systemic relapse of the disease is frequent despite complete resection. Five-year overall survival (OS) rates are reported to be 70% for stages IB-IIA, 50%–60% for stages IIAIIB, and 35% for stage IIIA (3). However, some patients experienced a cure for lung cancer by chest surgery alone. In contrast, other patients with completely resected NSCLC may have a minimal residual disease (MRD) not identified radiographically and require adjuvant chemotherapy.

Thus, adjuvant chemotherapy is indicated to avoid recurrence in patients who underwent complete surgical resection NSCLC. For patients with stage IIA–IIIA NSCLC postoperative, the standard treatment adjuvant is cisplatin-based chemotherapy. Pooled analysis of several clinical trials has shown that adjuvant cisplatin-doublet chemotherapy improved the 5-year disease-free survival (DFS) rate by 5.8% and the 5-year OS rate by 5.4% (4). Moreover, the ADAURA trial demonstrated that osimertinib improved DFS in epidermal growth factor receptor (EGFR) mutation-positive stage IB–IIIA disease following complete resection (5). Therefore, the establishment of the genomic profile and targeted therapies may lead to improvements in therapeutic strategies and the clinical outcomes of lung cancer patients.

The “classical” EGFR mutations, such as the L858R point mutation in exon 21 and the in-frame deletional mutation in exon 19, have been undoubtedly associated with target therapy response, while many other EGFR mutations are rarely detected in patients with NSCLC, and

information about their association with response to target therapy are still uncertain (6).

In cell culture, rare exon 20 mutations make transformed cells less responsive to target therapy, suggesting that other mechanisms probably contribute to primary resistance in metastatic NSCLC (7). For many of the rare mutations, the effect on responsiveness remains unknown.

Thus, it is of extreme importance for the clinical decision-making process to share information of patients harboring such mutations, The “classical” EGFR mutations, such as the L858R point mutation in exon 21 and the in-frame deletional mutation in exon 19, have been undoubtedly associated with target therapy response, while many other EGFR mutations are rarely detected in patients with NSCLC, and information about their association with response to target therapy are still uncertain (6). In cell culture, rare exon 20 mutations make transformed cells less responsive to target therapy, suggesting that other mechanisms probably contribute to primary resistance in metastatic NSCLC (7). For many of the rare mutations, the effect on responsiveness remains unknown. Thus, it is of extreme importance for the clinical decision-making process to share information of patients harboring such mutations, particularly when the outcome of epidermal growth factor receptor-tyrosine kinase receptor (EGFR-TKI) treatment is available.

Previous studies performed by our group using frozen tissue specimens and nextgeneration sequencing (NGS) from a cohort of 70 patients have shown that Brazilian patients represent a genetically admixed NSCLC cohort with important predictive and prognostic implications (8, 9). In the same case series and array used (QIAamp DNA Mini Kit Qiagen), we detected some Brazilian patients with NSCLC harboring rare EGFR variants with clinical significance to be defined.

Therefore, the aim of the present study was to explore the clinical outcome of patients with resected NSCLC in the early stage harboring rare somatic EGFR mutations. We also

considered mutations in exons 18 and 20, uncommon mutations in exons 19 and 21, and/or the presence of different mutations in a single tumor (complex mutations) were all considered rare. We found 14 cases with rare mutations. These patients were treatedwith adjuvant chemotherapy and two of them with erlotinib. In this scenario, we discuss their clinical outcome case by case with references to the literature if obtainable, or in silico analysis. We also highlight two rare pathogenic mutations that were identified and one case showing EGFR-TKI efficacy with such a mutation. The other rare mutations have been previously described in case reports.

Material and Methods

Sample Collection

Since that novel variant could also arise from sequencing artifacts, especially the ones associated to formalin related DNA damage as described in Wong, S.Q et al (9), our investigation was done in fresh-frozen specimens from Brazilian patients with lung cancer, collected during surgical resection from August 2003 to August 2010 at the A.C. Camargo Cancer Center, a tertiary referral center for the treatment of lung cancer in São Paulo, Brazil. The histologic classification was reviewed according to the World Health Organization 2021 classification system (11) and stratified as non-squamous non-small lung cancer (46 cases) and squamous non-small cell lung cancer (n=24). EGFR status was associated with baseline characteristics (age, sex, ethnicity, smoking history, stage at diagnosis, histology, type of EGFR mutation), treatments (radiotherapy, platinum-based chemotherapy, tyrosine kinase inhibitors) and outcome (relapse-free survival [RFS] after primary surgical resection, relapse, and development of distant metastases.

The study was approved in accordance with the ethical standards of the local committee on human experimentation (Research Ethics Committee of University of São Paulo Medical

School - CAAE: 79769017.1.0000.5440; opinion number: 2.673.320). Informed consent was waived because of the retrospective study design. The identity of the subjects under this retrospective analysis was also omitted and anonymized.

EGFR Mutation Analysis

Next-Generation Sequencing Analysis

The DNA of fresh tumor tissue was extracted using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's recommendations, and quantified using the Qubit® 3.0 Fluorometer (Invitrogen, Life Technologies, Carlsbad, CA, USA). The full genomic sequence of EGFR (2 kbp upstream 5'UTR, 2 kbp downstream 3'UTR, coding regions, and intronic sequences) was targeted and captured using TruSeq Custom Amplicon Panel v1.5kit (TSCAP, Illumina, SanDiego, CA) followed by massively parallel sequencing of enriched fragments on an Illumina MiSeq platform (Illumina, San Diego, CA, USA) consisting of 150 bp paired-end reads (300 cycles). All tumor specimens had an average sequencing depth of the target region ≥100× and coverage of the target region ≥90% at 30×.

Identification of Single Nucleotide Variants (SNVs) and Indels

In order to reducing the effects of PCR amplification and sequencing artifacts, sequencing data analyses was performed on the Molecular Genetics and Bioinformatics Laboratory of the Experimental Research Unit (UNIPEX) at the Medical School of São Paulo State University (UNESP). Sequencing data analyses were performed on the Molecular Genetics and Bioinformatics Laboratory of the Experimental Research Unit (UNIPEX) at the Medical School of São Paulo State University (UNESP). The raw sequencing data were base-called and demultiplexed using MiSeq Reporter v.1.8.1 (Illumina, San Diego, CA, USA) with default parameters, and FastQC files were generated for downstream data analysis. Filtered reads were aligned to the human genome (hg19, GRCh37) using the Burrows-Wheeler Alignment tool

(BWA) v.0.7.10. After alignment, the SAMtools software was applied to convert the alignment files to an indexed binary alignment map format. The single nucleotide variants (SNVs) and short insertions and deletions (INDELs) were called using the GATK UnifiedGenotyper, including HaplotypeCaller with default parameters based on hg19 and annotated with dbSNP version 144. To reduce false-positive somatic mutations which might originate from germline variants were followed the following cut off criteria: number of reads with the altered base in the tumor ≥10, mutations detected at a position of total read depth of ≥ 100, frequency of the reads with the altered base in the tumor ≥ 5% except for variants that are also reported in the COSMIC database, minor allele frequency < 0.1% in two publicly available databases, namely 1000 Genomes, and Exome Aggregation Consortium.

In silico Prediction Tools and Tumor Samples to Guidance Uncommon Profiles

The variants were annotated using the VEP software based on the consequences, predicted impacts, and reported allele frequencies in the population. Variants of unknown significance (VUS) were checked on the Cancer genome interpreter (https://www.cancergenomeinterpreter.org/home) and VarSome database, a single website that allows access to publications, ClinVar, and all in silico prediction tools including the Genomic Evolutionary Rate Profiling (GERP). We also used the Mastermind bioinformatic platform to examine further clinical diagnostic variant interpretation. Next, the mutation plots were generated using the online customized MutationMapper tool at cBioPortal for Cancer Genomics. In this study, all variants were classified according to the American College of Medical Genetics and Genomics (ACMG) recommendations, which allows classification as either "pathogenic," "likely pathogenic," "likely benign," "benign" or "uncertain significance".

Data evaluation

All analyses of the cohort were exploratory. No statistical test was applied given the small size of the cohort, implying a reduced power of significance.

Results

Characteristics of the study cohort

Table 1 summarizes the baseline characteristics of the cohort. 70 Brazilian patients with NSCLC surgically resected were included in the study. The median age of the patients was 78 years (range, 41–96), with similar distribution between older and younger patients. Forty-four patients were male (62.9%), 48 (68.6%) with European ancestrally (68.6%), and 21 (30.0%) smokers. Adenocarcinoma histotype was diagnosed in 46 (65.71%) patients and squamous cell carcinoma in 24 (34.3%) patients. Classification of patients according to TNM staging included 14 individuals (20.0%) in stage IA, 10 individuals (14.3%) in stage IB, 10 individuals (14.3%) in stage IIA, 13 individuals (18.6%) in stage IIB, and 11 individuals (15.7%) in stage IIIA. During the postoperative period, pulmonary and mediastinal relapse occurred in 11 patients (15.7%) and distant metastasis (brain and bone) in 12 patients (17.4%). Of the 70 patients, 23 (32.8%) received platinum-based chemotherapy, and two patients (2.9%) received erlotinib. Thirty-eight patients (64%) died. The median follow-up was 49 months (range, 0-175).

Table 1. Frequency of demographic and clinical characteristics of 70 NSCLC patients.

Characteristics	Number of Patients (N=70)
Age, years	
Mean (range)	77 (41-96)
≤77	32 (45.7%)
>77	33 (47.1%)
Gender	
Male	44 (62.9%)
Female	21 (30.0%)
Race	
White	48 (68.6%)
Asian	2 (2.9%)
Brown	2 (2.9%)
Smoking status	

Smoker	21 (30.0%)
Non-smoker	7 (10.0%)
Histological subtype	
Adenocarcinoma	46 (65.71%)
Squamous cell carcinoma	24 (34.3%)
Pathological TNM stage†	
IA	14 (20.0%)
IB	10 (14.3%)
IIA	10 (14.3%)
IIB	13 (18.6%)
IIIA	11 (15.7%)
Relapse	
No	13 (18.6%)
Locorregional	11(15.7%)
Distant metastasis	12 (17.4%)
Post operative treatment	
Chemotherapy platinum-based	23 (32.8%)
Radiotherapy	4 (4.3%)
Chemoradiotherapy	6 (8.6%)
Tyrosine kinase inhibitor (erlotinib)	2 (2.9%)
Status for overall survival	
Live	23 (32.9%)
Dead	38 (54.3%)
Follow-up (months)	49 (0-175)
***EGFR* mutational status**	
Classic Mutations (18-21 exons)	14 (20.0%)
Uncommon mutation	9 (12.9%)
Wild type	47 (67.1%)

Abbreviations: NSCLC, non-small cell lung cancer. Some cases lacked follow-up information: gender (5); age (5); race (18); smoking status (42); TNM stage (6); Relapse (36); Chemotherapy (47); Radiotherapy (43), Status (9). † 8th Edition International Association for the Study of Lung Cancer.

Frequency of EGFR somatic mutations

In our casuistry, EGFR mutations (target exons 1-28) were detected in 23 of 70 patients (32.9%), 44 males and 21 females. Classic or sensitizing activating mutations (exons 18-21) occurred in 14 patients (20.0%), being 9 of 14 (12.9%) in exon 19 (inframe deletion). Uncommon mutations in exons 3, 7, 14, 16, 22, 27, and 28, and/or complex mutations were identified in 9 tumors corresponding to 12.9% of the entire population analyzed. In **Table 2** and **Table 3**, the results of NGS sequences are listed in detail.

Table 2: Spectrum of *EGFR* variants in exons 18-21 identified in a Brazilian NSCLC cohort by Next-Generation Sequencing.

Exon/Intron	ID rs	HGVS Nucleotide	HGVS Protein	Variant Type	Frequency	ACMG Classification
18	rs28929495	c.2155G>T	G719C (p.Gly719Cys)	Missense	1	Pathogenic
19	rs121913229	c.2248G>C	A750P (p.Ala750Pro)	Missense	7	Likely Pathogenic
19	rs121913421	c.2235_2249del	E746_A750del (p.Glu746_Ala750del)	inframe_deletion	5	Pathogenic
19	rs121913436	c.2239_2247delTTAAGAGAA	L747_E749del (p.Leu747_Glu749del)	inframe_deletion	1	Likely Pathogenic
19	rs121913442	c.2240_2254del	L747_T751del (p.Leu747_Thr751del)	inframe_deletion	1	Likely Pathogenic
20	rs121913465	c.2303G>T	S768I (p.Ser768Ile)	Missense	1	Pathogenic
20	rs121434568	c.2438T>G	L813R (p.Leu813Arg)	Missense	2	Pathogenic
20	rs1275022697	c.2326C>T	R776C (p.Arg776Cys)	Missense	1	Likely Pathogenic
21	rs148934350	c.2543C>T	P848L (p.Pro848Leu)	Missense	1	Pathogenic

Abbreviations: EGFR, Epidermal Growth Factor Receptor; NSCLC, *Non-small Cell Lung Cancer*; rs, Reference Single Nucleotide Polymorphism; HGVS, Human Genome Variant Society; ACMG, American College of Medical Genetics and Genomics.

Table 3: Spectrum of uncommon EGFR variants identified in a Brazilian NSCLC cohort by Next-Generation Sequencing.

Exon/Intron	ID rs	HGVS Nucleotide	HGVS Protein	Variant Type	Frequency	ACMG Classification
3	rs754854319	c.409A>G	M137V (p.Met137Val)	Missense	3	Uncertain Significance
7	rs149840192	c.866C>T	A289V (p.Ala289Val)	Missense	1	Likely Pathogenic
8	rs886037891	c.977G>T	C326F (p.Cys326Phe)	Missense	1	Likely Pathogenic
14	rs144943614	c.1639G>A	V547I (p.Val547Ile)	Missense	1	Uncertain Significance
14	rs779076899	c.1705G>A	G569S (p.Gly569Ser)	Missense	1	Uncertain Significance
15	rs538888597	c.1881-462C>T	*	Intron	1	Uncertain Significance
16	rs369399038	c.1903C>T	R635W (p.Arg635Trp)	Missense	1	Uncertain Significance
22	rs376822837	c.2611G>A	D871N (p.Asp871Asn)	Missense	1	Uncertain Significance
27	rs35918369	c.3494C>T	A1165V (p.Ala1165Val)	Missense	1	Uncertain Significance
28	rs869064669	c.*582T>G	*	3_prime_UTR_variant	1	Uncertain Significance
28	rs751311059	c.*708T>C	*	3_prime_UTR_variant	1	Uncertain Significance
28	rs988454507	c.*934G>A	*	3_prime_UTR_variant	1	Uncertain Significance
28	rs771422383	c.*1060G>A	*	3_prime_UTR_variant	1	Uncertain Significance

Abbreviations: EGFR, Epidermal Growth Factor Receptor; NSCLC, *Non-small Cell Lung Cancer*; rs, Reference Single Nucleotide Polymorphism; HGVS, Human Genome Variant Society; ACMG, American College of Medical Genetics and Genomics.

Clinical outcome in patients harboring mutations cluster in exons 18-21

Exon 18 Mutations

G719C (exon 18) + S768I (exon 20)

Patient report. The G719 + S7681 pathogenic double mutation was found in a 70-yearold male smoker with European ancestrally and adenocarcinoma showing micropapillary characteristics in stage IIIA. After local recurrence, the patient was treated with a combination chemotherapy of CDDP and etoposide. The patient died 13 months after the surgical resection.

Reference. The complex mutation has been previously described in a few reports (12,13). In vitro data suggest that the coexistence of the E709A + G719C mutation confers resistance to adjuvant treatment, including EGFR-TKIs (14). Another report did not confirm the low sensitivity to adjuvant treatment of the double mutant (15).

Comment. According to the cases described so far, the G719C + S768I complex mutation seems to be associated with resistance to adjuvant treatment including EGFRTKI.

Exon 19 Mutations

p.E746_R748del/p.A750P

Patient report. The p.E746_R748del/p.A750P pathogenic complex mutation was found in two patients female nonsmokers patients, 90-year-old and 96-year-old, both with European ancestrally, and adenocarcinoma stage IIIA with acinar and solid characteristics. After pulmonary and mediastinal local relapse, both patients were treated with carboplatin and Taxol, which was interrupted early due to toxicity. A clinically meaningful improvement of symptoms was obtained within a few days and a partial response of tumor assessment after 6 weeks. After 9 months, both patients presented disease progression and died.

Reference. The presence of p.E746_R748del/p.A750P has already been reported in one Caucasian patient by Jakobsen and colleagues (16) but without specific information on the efficacy of EGFR-TKI.

Comment. Similarly, from the single case reported in the literature showing progression following gefitinib, the two patients treated in our institution had a rapid symptomatic improvement but died 9 months after disease progression.

Exon 20 Mutations

L813R + R776C

Patient report. The L813R + R776C EGFR mutation was found in a 70-year-old current heavy male smoker, with European ancestrally and adenocarcinoma stage IIIA with acinar characteristics. After pulmonary local relapse, the patient was treated with CDDP/Navelbine. A tumor progression was observed after 10 weeks at the first treatment assessment. Erlotinib was subsequently given at the starting dose of 150 mg daily. The patient experienced 30 months of progression-free survival.

Reference. The L813R + R776C mutation was previously described by Ham and colleagues (17) in a patient with squamous NSCLC, who benefited from afatinib and experienced 11 months of progression-free survival. The same mutation has been found by Moran and colleagues (18), but in association with the L861 in exon 20, in a neversmoking patient with adenocarcinoma with lepidic features. In this case, the patient had a complete and long-lasting remission after afatinib treatment.

Comment. Similarly, from the case reported by Moran where the L861 was linked to R776C, in our experience the L813R + R776C EGFR mutation was associated with response to erlotinib.

Exon 21 Mutations

P848L

Patient report. The P848L mutation was found in a 94-year-old current heavy male smoker, with European ancestrally and squamous cell carcinoma with non-keratinizing features. After

lobectomy, the tumor pathological stage was classified as IB. During the first three months of follow-up, the patient developed pulmonary relapse and boné metastases. He was treated with radiotherapy and a combination of CDDP/Navelbine, followed by erlotinib with rapid tumor progression, and died.

Reference. P848L has been described in 16 carcinoma patients in the Catalogue of Somatic Mutations in Cancer (COSMIC) (19), representing a frequency of about 0.6% in cancer patients (20). A recent study showed that the P848L germline variant of EGFR contributes to cellular transformation and predicts resistance to TKIs, and patients harboring P848L likely benefitting from JAK inhibitors (21). Furthermore, very rare exon 21 mutations have shown to be more sensitive to afatinib (second-generation EGFRTKIs) and osimertinib (third-generation EGFR-TKIs) than first-generation EGFR-TKIs in preclinical studies (22).

Comment. This is an unusual case of a patient with a tumor in stage IB harboring a P848L mutation with no response to adjuvant treatment, including TKI, and rapid progression of the disease.

Clinical outcome of patients harboring rare EGFR mutations with uncertain significance

Exon 3 Mutations

M137V

Patient report. M137V mutation was found concurrently with other two unknown mutations, c.1881-462C>T (intron15) and A1165V (exon 27), in a 78-year-old female non-smoker with European ancestrally and adenocarcinoma showing lepidic characteristics. The patient underwent pneumectomy and the tumor was staged as IB. She received adjuvant radiotherapy and a combination of CDDP with Navelbine with partial response of the tumor and is alive after 8 years of follow-up.

Reference. The M137V mutation and the c.1881-462C>T and A1165V variants are reported in the dbSNP database (rs754854319) without any data regarding their clinical significance.

Chakroborty et al. (2019) described this alteration in colon adenocarcinomas from human samples available in the GENIE database (23). VarSome database predicted their significance to be uncertain. The DEOGEN2, FATHMM-MKL, LIST-S2, M-CAP, MutationAssessor, MutationTaster, and SIFT computational analysis inferred pathogenicity, while BayesDel_addAF, DANN, EIGEN, MVP, and PrimateAI tools) predicted benignity. The GERP score (5.13) indicates that the Met amino acid in position 137 lies in a strongly conserved region, further deposing in favor of predicting pathogenicity.

Comment. Our patient with adenocarcinoma stage IB harboring M137V mutation responded partially to adjuvant treatment and is alive after 8 years of follow-up.

Exon 7

A289V Mutation

Patient report. The A289V mutation was found in a 94-year-old male patient, with European ancestrally and adenocarcinoma with solid characteristics, and staged as IIB. The patient died two months after the surgical resection.

Reference. The A289V mutation in exon 7 encodes an amino acid of the extracelular domain of the EGFR protein and had been found only in the glioblastoma, low-grade glioma, head, and neck neoplasms, being rarely described in NSCLCs (24, 25) . Dai et al. (2018) were the first to present a case report of NSCLC in a Chinese patient with EGFR p.A289V mutation, whose corresponding treatment showed only a partial response to first-generation EGFR-TKIs (25). The A289V mutation was classified as likely pathogenic by ACMG. VarSome database analysis showed a pathogenic computational result based on ten pathogenic predictions from BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MutationAssessor, MutationTaster, and SIFT vs. two benign predictions from MVP and PrimateAI.

Comment. The A289V mutation seems to occur in patients with advanced age in the early

stage of the disease rapidly progressing to death.

Exon 8.

C326F Mutation

Patient report. C326F mutation was detected in a stage IIB keratinizing squamous cell carcinoma from a smoker patient 89-year-old, who underwent pneumonectomy and died one month after.

Reference. This genomic alteration had been described in a case report of Lhermitte- Duclos disease, also known as dysplastic cerebellar gangliocytoma, which has been shown to be an activation mutation resulting in increased EGFR autophosphorylation and indicating high receptor activation (26). VarSome's in silico analysis of the C326 mutation revealed 11 pathogenic predictions from BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MVP, MutationAssessor, MutationTaster, and SIFT vs. one benign prediction from PrimateAI.

Comment. The C326F also occurred in a patient with advanced age with squamous cell carcinoma in an early atage disease rapidly progressing to death.

Exon 14

V547I-G569S Mutation

Patient report. The V5471-G569S mutation, concomitantly with a point mutation R776C, was found in a 69-year-old male patient with African ancestrally and nonkeratinizing squamous cell carcinoma, stage IIB, who presented pulmonary relapse 11 months after surgical resection. The patient was treated with radiotherapy and a combination of carboplatin and Taxol but died due to brain metastases.

Reference. VarSome database analysis revealed seven pathogenic predictions for the V5471-G569S mutation according to DANN, EIGEN, EIGEN PC, FATHMM-MKL, FATHMM-XF,

LRT, MutationAssessor, and MutationTaster against 12 benign predictions from BayesDel addAF, BayesDel noAF, DEOGEN2, FATHMM, LIST-S2, MVP, MetaLR, MetaSVM, PrimateAI, REVEL, and SHIFT 4G. The G569S mutation with concomitant point mutation R776C in the VarSome computational analysis led to a benign computational result based on 12 benign predictions from BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MVP, MutationAssessor, MutationTaster, PrimateAI and SIFT vs. no pathogenic predictions.

Comment. In our case, the V5471-G569S mutations were detected in non-keratinizing squamous cell carcinoma, stage IIB, from a patient of African ancestry who did not respond to adjuvant chemotherapy and died after brain metastases.

Exon 16

R635W Mutation

Patient report. The mutation R635W was found in a 64-year-old female non-smoker patient with adenocarcinoma, stage IIA, who received adjuvant chemotherapy and radiotherapy and died three years after the surgery.

Reference. The mutation R635W is reported in the dbSNP database (rs369399038) of uncertain significance. In silico analysis predicted a deleterious effect based on 11 pathogenic predictions from BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMMMKL,

LIST-S2, M-CAP, MutationAssessor, MutationTaster, PrimateAI, and SIFT vs. one benign prediction from MVP.

Comment. Another patient from our Institution with adenocarcinoma stage IIA harboring the mutation R635W experienced partial response to radiotherapy and chemotherapy with rapid progression of the disease.

Exon 22

D871N Mutation

Patient report. The D871N mutation was found in a 69-year-old heavy smoker male patient with African ancestrally and keratinizing squamous cell carcinoma, stage IIB, who underwent pneumectomy. The tumor relapsed in the central nervous system. The patient received a combination of carboplatin and Taxol and died eleven months after.

Reference. The VarSome database produced a final likely pathogenic computational result based on seven pathogenic predictions from DANN, DEOGEN2, FATHMM-MKL, LIST-S2, M-CAP, MutationTaster, and SIFT vs. five benign predictions from BayesDel_addAF, EIGEN, MVP, MutationAssessor, and PrimateAI.

Comment. Another uncommon mutation harbored in squamous cell carcinoma from Patient with poor tumor response.

Exon 28 and Exon 29

Non-coding mutations

Patient report. Three male patients older than 70 years with IIA adenocarcinoma harboring respectively c.*582T>G (rs869064669), c.*708T>C (rs751311059), c.*934G>A (rs988454507), and c.*1060G>A (rs771422383) who underwent lobectomy and received a combination of CDDP with Navelbine are alive nine years after surgical resection.

Reference. The four VarSome analyses resulted in benign computational results, with one benign prediction from DANN vs. no pathogenic prediction in all four cases. Moreover, all four variants had a GERP score of less than 5.5, an indication that the position of these variants was not highly conserved (-0.878, 3.56, -2.5, and -0.396, respectively).

Comment. It is worth noting, however, that the use of more sensitive sequencing techniques might reveal rare mutants at a higher frequency.

Discussion

Previous studies from the present case series have shown that Brazilian patients represent a genetically admixed NSCLC cohort with clinical implications. In fact, we found that patients with adenocarcinoma harboring somatic mutations of EGFR, CTLA4, PDCD1LG2, or ZEB2 that only underwent surgical treatment and developed brain metastases may have the worst prognosis (8,9). Regarding EGFR, we detected approximately 13% of NSCLC Brazilian patients harboring EGFR somatic mutations with clinical significance to be defined. These findings motivated us to re-analyze EGFR mutations in the 70-case series of NSCLC. Because of the small sample size, we discuss their clinical outcome case by case with references to the literature, and not based on powered for significance.

Under the objectives of the present study, we found: (1) rare mutations, uncommon mutations, and/or complex somatic mutations identified in 20% of Brazilian population analyzed with prevalence of European ancestrally, non-smoker male older than 77 yrs, adenocarcinoma histology stage IB-IIB; (2) rare G719C (exon 18) + S768I (exon 20) complex somatic mutation seems to be associated with resistance to combination chemotherapy of CDDP and etoposide with short PFS; (3) rare p.E746_R748del/p.A750P complex somatic mutations resistant to carboplatin and Taxol with short PFS; (4) rare L813R + R776C (exon 20) complex somatic mutation associated with response to erlotinib and prolonged PFS; (5) rare P848L (exon 21) somatic mutation resistant to combination of CDDP/Navelbine and erlotinib with short PFS; (6) uncommon M137V (exon 3) and non-coding (exons 28 and 29) somatic mutations with response to combination of CDDP with Navelbine and prolonged PFS; (7) uncommon somatic mutations A289V (exon 7), C326F (exon 8), V547I-G569S (exon 14), R635W (exon 16), D871N (exon 22) with resistance to combination of carboplatin and Taxol and short PFS.

The frequency of EGFR mutations in NSCLC ranging from 5 to 20%, depending on the cohort studied (6). Erlotinib and gefitinib are small molecules that target tumor EGFR mutation

and are effective to treat patients with this subgroup of tumors, showing a response rate of approximately 70% (27). Two classical activating mutations represent the most common mutations: a short in-frame deletion of exon 19 and a point mutation (CTG to CGG) in exon 21 at nucleotide 2573 leading to the replacement of leucine by arginine at codon 858 (L858R). However, other EGFR mutations do not effectively cause EGFR tyrosine-kinase activity and may be associated with acquired TKIs resistance, therefore affecting the sensitivity of the EGFR-TKIs (28). In the case of rare EGFR mutations, the literature is very limited in terms of clinical outcome and response to treatment and may underpower the value of target therapy.

Therefore, from an EGFR mutation screen, performed on our patients, we detected NSCLC with rare EGFR mutations and illustrated the response rate after chemotherapy and erlotinib treatment. In detail, we analyzed 70 tumors and identified 14 cases, with rare mutations, defined as mutations in exons 18, 20, and 21 uncommon mutations in exons 3, 7, 14, 16, 22, 27, 28, and/or complex mutations (different mutations present in a single tumor). These mutations accounted for approximately 20% of all the cases screened coinciding with other studies in South America including Brazil (29) and Colombia (30) but was lower than those found in Asia and higher than those of European populations (31, 32). The current cohort included patients from all regions of the country who came to a tertiary referral center for the treatment of lung cancer. These patients present different education degrees, socioeconomic status and ancestrally characterizing a genetic admixture, inherited from European, African, and Asian immigrants (31, 32). As expected, we observed a high prevalence of rare and uncommon EGFR mutation in Brazilian patients with European ancestry, compared to those found in the native European population (33). Twenty-three of our patients were treated with platin-based chemotherapy and two patients received erlotinib, and we report herein the individual case reports regarding the literature and in silico analysis. We observed that patients with rare EGFR mutations could receive platinum-based chemotherapy as a first-line treatment, due to their low response rates and short PFS in response to erlotinib. The clinical outcome observed supported

the different predictive values of the single EGFR mutations in terms of treatment effectiveness. Of note, in our cohort, exon 20 mutations were associated with erlotinib sensitivity, and the patient experienced 30 months of progression-free survival; however, there are only a few reports in the literature of cases responding to EGFR-TKIs. Therefore, we infer that the evaluation of the single mutations, case by case, could be useful also in the presence of exon 20 mutations, as in the presence of other uncommon EGFR mutattions.

Although we showed the potential of NGS technique in small cohort of NSCLC, future validation is necessary using a similar cohort with a large set of patients to corroborate the results observed in this tumor tissue type. The main limitation for our exploratory analysis is the small number of NSCLC cases used but it was minimized by the data obtained using this technique. The present study is largely descriptive and exploratory, and extension of our findings are essential. Another limitation of our research was that the comprehensive interpretation of sequence variants requires not only wellestablished databases but also appropriate functional analyses, such as evaluating the effect of gene expression of these found variants.

In summary, we have provided complementary information concerning the effectiveness of chemotherapy and erlotinib in patients with NSCLC early stage for several rare EGFR mutations not previously or only rarely reported. This data together with similar studies in the literature and in silico analysis reinforce the decision-making process in such subsets of patients.

References

1. de Sá VK, Coelho JC, Capelozzi VL, de Azevedo SJ. Lung cancer in Brazil: epidemiology and treatment challenges. Lung Cancer (Auckl) 2016 Nov 14;7:141-148.doi: 10.2147/LCTT.S93604.

2. Araujo LH, Baldotto C, Castro G Jr, Katz A, Ferreira CG, Mathias C, et al. Lung câncer in Brazil. J Bras Pneumol 2018 Jan-Feb;44(1):55-64.doi: 10.1590/S1806-37562017000000135.

3. Monteiro AS, Araújo SRC, Araujo LH, Souza MC. Impact of microvascular invasion on 5-year overall survival of resected non-small cell lung cancer. J Bras Pneumol 2022 Jul 8;48(3):e20210283. doi: 10.36416/1806-3756/e20210283.

4. Booth CM, Shepherd FA, Peng Y, Darling G, Li G, Kong W, et al. Adjuvant chemotherapy for non-small cell lung cancer: practice patterns and outcomes in the general population of Ontario, Canada. J Thorac Oncol 2012 Mar;7(3):559-66. doi:10.1097/JTO.0b013e31823f43af

5. Zhang SS, Ou SI. Deconstructing ADAURA: It is Time to Forgo Adjuvant Platinum- Based Chemotherapy in Resected IB-IIIA EGFR+ NSCLC (Except with RB Alterations?) When Adopting Adjuvant Osimertinib. Lung Cancer (Auckl) 2022 Apr 26;13:23-31. doi:10.2147/LCTT.S358902.

6. Riely GJ, Pao W, Pham D, Li AR, Rizvi N, Venkatraman ES, et al. Clinical course of patients with non-small cell lung cancer and epidermal growth factor receptor exon 19 and exon 21 mutations treated with gefitinib or erlotinib. Clin Cancer Res 2006 Feb1;12(3 Pt 1):839-44. doi: 10.1158/1078-0432.CCR-05-1846.

7. Greulich H, Chen TH, Feng W, Jänne PA, Alvarez JV, Zappaterra M, et al. Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. PLoS Med 2005 Nov;2(11): e313. doi: 10.1371/journal.pmed.0020313.

8. Parra ER, Jiang M, Machado-Rugolo J, Yaegashi LB, Prieto T, Farhat C, et al. Variants in Epithelial-Mesenchymal Transition and Immune Checkpoint Genes Are Associated With Immune Cell Profiles and Predict Survival in Non-Small Cell Lung Cancer. Arch Pathol Lab Med 2020 Oct 1;144(10):1234-1244. doi: 10.5858/arpa.2019-0419-OA.

9. Machado-Rugolo J, Gutierrez Prieto T, Fabro AT, Parra Cuentas ER, Sá VK, Baldavira CM, et al. Relevance of PD-L1 Non-Coding Polymorphisms on the Prognosis of a Genetically Admixed NSCLC Cohort. Pharmgenomics Pers Med 2021 Feb 15; 14:239-252. doi: 10.2147/PGPM.S286717.

10. Wong SQ, Li J, Tan AY, Vedururu R, Pang JM, Do H, Ellul J, Doig K, Bell A, MacArthur GA, Fox SB, Thomas DM, Fellowes A, Parisot JP, Dobrovic A; CANCER 2015 Cohort. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. BMC Med Genomics. 2014 May 13;7:23. doi: 10.1186/1755-8794-7-23.

11. Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, et al. The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. JThorac Oncol 2022;17(3):362-387. doi: 10.1016/j.jtho.2021.11.003.

12. Chantharasamee J, Poungvarin N, Danchaivijitr P, Techawatanawanna S. Clinical outcome of treatment of metastatic non-small cell lung cancer in patients harboring uncommon EGFR mutation. BMC Cancer 2019 Jul 17;19(1):701. doi: 10.1186/s12885-019-5913-9.

13. Bi H, Ren D, Wu J, Ding X, Guo C, Miura S, et al. Lung squamous cell carcinoma with rare epidermal growth factor receptor mutation G719X: a case report and literature review. Ann Transl Med 2021 Dec;9(24):1805. doi: 10.21037/atm-21-6653.

14. Tam IY, Leung EL, Tin VP, Chua DT, Sihoe AD, Cheng LC, et al. Double EGFR mutants containing rare EGFR mutant types show reduced in vitro response to gefitinib compared with common activating missense mutations. Mol Cancer Ther 2009 Aug;8(8):2142-51. doi: 10.1158/1535-7163.MCT-08-1219.

15. Hijiya N, Miyawaki M, Kawahara K, Akamine S, Tsuji K, Kadota J, et al. Fosphorylation status of epidermal growth factor receptor is closely associated with responsiveness to gefitinib in pulmonary adenocarcinoma. Hum Pathol 2008 Mar;39(3):316-23. doi: 10.1016/j.humpath.2007.10.027.

16. Jan Nyrop Jakobsen, Eric Santoni-Rugiu, Morten Grauslund, Linea Melchior, Jens Benn

Sørensen. Concomitant driver mutations in advanced EGFR-mutated non-smallcell lung cancer and their impact on erlotinib treatment. Oncotarget 2018 May 25; 9(40):26195–26208.

17. Han C, Ding X, Li M, Luo N, Qi Y, Wang C. Afatinib, an effective treatment for patient with lung squamous cell carcinoma harboring uncommon EGFR G719A and R776C co-mutations. J Cancer Res Clin Oncol 2022 May;148(5):1265-1268. doi:10.1007/s00432-021-03864-4.

18. Moran T, Taus A, Arriola E, Aguado C, Dómine M, Rueda AG, et al.; Study Group for the Uncommon EGFR Mutations in Spain. Clinical Activity of Afatinib in Patients With Non-Small-Cell Lung Cancer Harboring Uncommon EGFR Mutations: A Spanish Retrospective Multicenter Study. Clin Lung Cancer 2020 Sep;21(5):428-436.e2. doi: 10.1016/j.cllc.2020.04.011.

19. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: theCatalogue Of Somatic Mutations In Cancer. Nucleic Acids Res 2019; 47: D941-D947.doi: 10.1093/nar/gky1015.

20. Santamaría I, Menéndez ST, Balbín M. EGFR L858R mutation may go undetected because of P848L in cis mutation. J Clin Oncol 2013; 31: e420-1. doi: 10.1200/JCO.2012.47.3512.

21. Sarcar B, Gimbrone NT, Wright G, Remsing Rix LL, Gordian ER, Rix U, et al. Characterization of epidermal growth factor receptor (EGFR) P848L, an unusual EGFR variant present in lung cancer patients, in a murine Ba/F3 model. FEBS Open Bio 2019; 9:1689-1704. doi: 10.1002/2211-5463.12702.

22. Zhang H. Three generations of epidermal growth factor receptor tyrosine kinase inhibitors developed to revolutionize the therapy of lung cancer. Drug Des Devel Ther 2016; 10:3867-3872. doi: 10.2147/DDDT.S119162.

23. Chakroborty D, Kurppa KJ, Paatero I, Ojala VK, Koivu M, Tamirat MZ, et al. An unbiased in vitro screen for activating epidermal growth factor receptor mutations. J Biol Chem 2019; 294:9377-9389. doi: 10.1074/jbc.RA118.006336.

24. An Z, Aksoy O, Zheng T, Fan QW, Weiss WA. Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies. Oncogene 2018; 37:1561-1575. doi: 10.1038/s41388-017-0045-7.

25. Dai L, Su X, Lu L, Lv D. Nonsmall cell lung cancer with rare exon 7 p.A289V mutation in the EGFR gene responds to Icotinib treatment: A case report. Medicine (Baltimore) 2018; 97: e13809. doi: 10.1097/MD.0000000000013809.

26. Colby S, Yehia L, Niazi F, Chen J, Ni Y, Mester JL, et al. Exome sequencing reveals germline gain-of-function EGFR mutation in an adult with Lhermitte-Duclos disease. Cold Spring Harb Mol Case Study 2016; 2: a001230. doi: 10.1101/mcs.a001230.

27. Jackman DM, Holmes AJ, Lindeman N, Wen PY, Kesari S, Borras AM, et al. Response and resistance in a non-small-cell lung cancer patient with an epidermal growth factor receptor mutation and leptomeningeal metastases treated with high-dose gefitinib. J Clin Oncol 2006 Sep 20;24(27):4517-20. doi: 10.1200/JCO.2006.06.6126.

28. Balak MN, Gong Y, Riely GJ, Somwar R, Li AR, Zakowski MF, et al. Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. Clin Cancer Res 2006 Nov 1;12(21):6494-501. doi: 10.1158/1078-0432.CCR-06-1570.

29. Freitas HC, Torrezan GT, da Cunha IW, Macedo MP, Karen de Sá V, Corassa M, et al. Mutational Portrait of Lung Adenocarcinoma in Brazilian Patients: Past, Present, and Future of Molecular Profiling in the Clinic. Front Oncol 2020;10:1068. doi:10.3389/fonc.2020.01068.

30. Arrieta O, Cardona AF, Corrales L, Campos-Parra AD, Sánchez-Reyes R, Amieva-Rivera E, et al.; CLICaP. The impact of common and rare EGFR mutations in response to EGFR tyrosine kinase inhibitors and platinum-based chemotherapy in patients with non-small cell lung cancer. Lung Cancer 2015 Feb;87(2):169-75. doi:10.1016/j.lungcan.2014.12.009.

31. Rahman S, Kondo N, Yoneda K, Takuwa T, Hashimoto M, Orui H, et al. Frequency of epidermal growth factor receptor mutations in Bangladeshi patients with adenocarcinoma of the lung. Int J Clin Oncol 2014; 19(1):45-9. doi: 10.1007/s10147-012-0515-4.

32. Smits AJ, Kummer JA, Hinrichs JW, Herder GJ, Scheidel‐Jacobse KC, Jiwa NM, et al. EGFR and KRAS mutations in lung carcinomas in the Dutch population: Increased EGFR mutation frequency in malignant pleural effusion of lung adenocarcinoma. Cell Oncol (Dodr) 2012; 35: 189‐196. doi: 10.1007/s13402-012-0078-4.

33. De Pas T Toffalorio F, Manzotti M, Fumagalli C, Spitaleri G, Catania C, et al. Activity of epidermal growth factor receptor-tyrosine kinase inhibitors in patients with non-small cell lung cancer harboring rare epidermal growth factor receptor mutations. J Thorac Oncol 2011 Nov;6(11):1895-901. doi: 10.1097/JTO.0b013e318227e8c614.

6. RESUMO E DISCUSSÃO DA TESE

O objetivo central desta tese foi promover o entendimento das vias de ativação e regulação das respostas celulares frente aos insultos e o desenvolvimento dos mecanismos subjacentes no reparo das lesões na patogênese do câncer de pulmão. A caracterização molecular das vias e fatores envolvidos nestas respostas ampliou a capacidade de identificar potenciais biomarcadores, com o intuito de colaborar no desenvolvimento de tratamentos mais eficazes e para o manejo clínico do câncer de pulmão. Um dos propósitos maiores embutidos nesta tese consistiu no aspecto translacional envolvendo investigações no câncer de pulmão atendendo ao seguinte objetivo específico: traçar o perfil de alteração genética do processo de EMT, analisando os genes relacionados ao desenvolvimento, à migração, a motilidade e à diferenciação celular, bem como os genes que regem as principais vias de sinalização envolvidas no evento da EMT em carcinomas de pulmão de células não pequenas. Esta tese foi o ponto inicial do Laboratório de Patologia e Genômica Pulmonar na utilização de Sequenciamento de Nova Geração (do inglês, *Next-Generation Sequencing*), também conhecido como Sequenciamento Massivo em Paralelo (*Massively Parallel Sequencing*) ou mesmo Sequenciamento de Alta Performance (do inglês, *High-throughput Sequencing*), conseguindo integrar toda a expertise do Laboratório em técnicas avançadas de imunohistoquímica com dados gerados de NGS.

Na era da medicina personalizada, a classificação histológica e a caracterização molecular são fundamentais para a decisão terapêutica. A evolução do conhecimento da biologia e genética molecular abriu caminhos para a criação de novos alvos terapêuticos e biomarcadores, acarretando um desafio no diagnóstico, tratamento e monitoramento das doenças. A informação embutida nessas áreas, particularmente no palco da regulação da expressão gênica e proteica, tem permitido o desenvolvimento de sistemas de diagnóstico eficazes para a identificação das mutações ou alterações genéticas responsáveis por uma variedade de doenças, assim como a sua utilização como marcadores prognósticos e utilização na escolha da melhor conduta terapêutica. No entanto, a medicina de precisão ainda está voltada para uma pequena parcela da população, principalmente no Brasil. A maioria dos pacientes com câncer de pulmão não são testados para mutações do receptor do fator de crescimento epidérmico (EGFR) e a maioria desses testes é realizada em laboratórios privados. A maior experiência no rastreamento da mutação do EGFR, por exemplo, em pacientes com CPNPC da rede pública de saúde é realizada em um programa de acesso patrocinado pela indústria no Brasil. No entanto, a indústria não fornece os medicamentos. A coorte estudada foi composta por 70 amostras de tecido tumoral provenientes de pacientes diagnosticados com CPNPC, os quais foram tratados e

acompanhados no Hospital da Clínicas de São Paulo (HCFMUSP), Instituto do Câncer de São Paulo (ICESP), Instituto do Coração (Incor) e A.C. Camargo Cancer Center. A maioria dos pacientes apresentou o subtipo histológico adenocarcinoma e recebeu como tratamento pós-operatório radioterapia e quimioterapia. Nesse sentido, nosso propósito maior em câncer de pulmão foi estudar os biomarcadores moleculares em tecidos, gerando evidências para pesquisas adicionais sobre novos medicamentos no tratamento de carcinomas de pulmão de células não pequenas.

Genericamente, o comprometimento das vias sinalizadoras de lesão pulmonar parenquimatosa, epitelial e matriz extracelular, envolve três estágios básicos: 1) iniciação como resultado de eventos intrínsecos e/ou extrínsecos para produzir disfunção celular; 2) amplificação da resposta imunocelular resultante e 3) resposta em que as células epiteliais passam por uma decisão de reparar apropriadamente via ativação de programas proliferativos e de diferenciação ou sofrer remodelamento inapropriado incluindo apoptose excessiva, transição epitélio-mesenquimal (EMT) ou reprogramação para um estado defeituoso e disfuncional de diferenciação. Este é o ponto de ramificação crítico no qual a correção do epitélio danificado ou anormalmente reparado ocorre, resultando no desenvolvimento de fibrose e / ou diferenciação para produzir uma neoplasia (64).

No estágio de iniciação, as alterações genéticas e epigenéticas celulares ocorrem após inflamação crônica e exposição ao tabagismo, incluindo ativação oncogênica e inativação do gene supressor devido ao dano ao DNA induzido pelo estresse oxidativo. Através da morte celular programada e proliferação celular, essas mudanças levam ao desenvolvimento de vários estágios e progressão incluindo metaplasia, hiperplasia adenomatosa atípica, displasia e podem chegar ao câncer de pulmão. Vários estudos relataram alterações em expressões genéticas e anormalidades moleculares envolvidas em vias de sinalização. As vias sinalizadoras mais alteradas são aquelas relacionada ao EGFR (EGF-R/Ras/PI3K/AKT/PTEN), P53 e Rb, apoptose e envelhecimento, parada G2 e reparo do DNA. A regulação da expressão gênica, sem modificações na sequência do DNA, por mecanismos moleculares caracterizam as alterações epigenéticas, que incluem: 1) alteração do estatus de metilação do DNA dentro de ilhas CpG, silenciamento de genes supressores tumorais em caso de hipermetilação ou ativação de oncogenes no caso de hipometilação; 2) modificação covalente das extremidades de histonas indutoras de remodelamento da cromatina; 3) microRNAs ligantes a 3'UTR do mRNA prevenindo a expressão proteica. Mais de 100 genes têm sido envolvidos no silenciamento de genes supressores tumorais com papel na progressão neoplásica incluindo RARβ, TIMP-3, P16, RASSF1, MGMT, FHIT, DAPK, E-caderina e GST1. Modificações pós-transcricionais na extremidade de histonas nucleossomais, como a trimetil histona H4 K20 emergem agora como

importantes processos sinalizadores epigenéticos ao controlar empacotamento da cromatina, expressão gênica e a estabilidade do genoma e provavelmente estão no desenvolvimento neoplásico. Alguns agentes moleculares direcionados surgiram para a inibição de alvos específicos entre as vias de sinalização desreguladas nas doenças respiratórias (65).

Após o estágio de iniciação celular ocorre a amplificação da resposta imunocelular inata e adquirida no microambiente tecidual. O sistema imune inato consiste de macrófagos, neutrófilos, células dendríticas [DCs], mastócitos, eosinófilos, basófilos, células assassinas naturais (NK) e células T NK. O sistema imunológico inato é a primeira linha de defesa contra células transformadas. Existem duas vias principais para a ativação de macrófagos in vivo, M1 ou M2. Normalmente, os macrófagos M1 são efetores imunológicos potentes, enquanto os macrófagos M2 suprimem as respostas imunes, promovem o remodelamento e angiogênese produzindo crescimento e fatores de sobrevivência para células transformadas (66). Macrófagos e mastócitos liberam mediadores tais como citocinas, quimiocinas, metaloproteinases de matriz, espécies reativas de oxigênio e mediadores bioativos (Histamina), que mobilizam leucócitos adicionais, angiogênese e remodelamento (67). As células NK (CD56+ CD-) desempenham um papel na rejeição de células tumorais (68). As células T NK (CD56+ CD3+) representam um subconjunto de linfócitos T, mas expressam uma variedade de marcadores moleculares típicos para células NK. Uma resposta imune inata leva à ativação do Sistema imunológico adaptativo (células B e T), Interação com células apresentadoras de antígenos e mediadores pró-inflamatórios. Os dois maiores subconjuntos de linfócitos T são: (1) células T auxiliares (Th, CD4) e (2) células T citotóxicas (CTL CD8+). Os linfócitos expressam receptores antigênicos somáticos específicos, permitindo um repertório flexível das respostas. Diferentes alterações genéticas e epigenéticas induzem resistência no sistema imune pela expressão de moléculas que se ligam com receptor para induzir o bloqueio do sistema imune e causar anergia ou exaustão dessas células. O maior ponto de checagem imune é o ligante de morte celular programada 1 (PD-L1), proteína que pode mediar a supressão da resposta imune contra os tumores. PD-L1 é uma glicoproteína de membrana celular que pertence à família das moléculas co-estimuladoras B7 assim como B7-H3 e B7-H4 tendo seu gene no cromossomo 9p24.2.42 (68).

O estágio de remodelamento ocorre sobretudo através da transição epitélio-mesenquimal (EMT). A EMT é o processo pelo qual as células epiteliais passam a expressar genes específicos de células mesenquimais, adquirindo, assim, fenótipo semelhante aos das células de mesênquima, e envolvida no remodelamento por fibrose dos órgãos (69), formação de tumores e metástases (70). A ativação da EMT é classicamente caracterizada por redução na produção

de E-caderina, uma molécula de adesão responsável pela aderência celular epitelial. Um segundo componente de complexos aderidos é a b-catenina, que desempenha parte integrante das junções aderentes, ligando a E-caderina ao citoesqueleto. Considerando-se as alterações epigenéticas descritas no câncer de pulmão, tem se observado que o gene CDH1, que codifica a E-caderina, encontra-se frequentemente inibido pela hipermetilação do seu promotor. As moléculas de N-caderina aumentam a afinidade das células neoplásicas pelas células estromais, permitindo a invasão. Fatores de crescimento e outras moléculas podem estar envolvidos nessa alteração de fenótipo, como o HGF (hepatocyte growth factor), o EGF (epidermal growth factor), o PDGF (platelet-derived growth factor), o VEGF (vascular endothelial growth factor) e o TGF-β (transforming growth factor β). Essas moléculas orquestram o evento, propiciando que as células neoplásicas se disseminem para locais distantes (71). Para tanto, após a perda de expressão de E-caderina, as células ganham motilidade e adquirem aspecto fibroblástico; estas células transformadas iniciam o processo migratório pela matriz extracelular (MEC). O remodelamento da MEC envolve grandes sistemas enzimáticos, proteínas estruturais e funcionais, interações bioquímicas e fisio-químicas, além de mediadores e respostas imunológicas (72). Neste contexto, o câncer de pulmão inclui-se entre as doenças respiratórias caracterizadas por iniciação celular, ampliação da resposta imune e remodelamento através da ativação da EMT.

Previamente a este estudo, avaliamos a expressão proteica e gênica de EGFR, P53, KRAS, ALK, ERBB2, PTEN, BRAF, VEGF, CD24 e CD44 (73). A imunoexpressão proteica de EGFR, EBB2, P53 e BRAF foi significativamente menor no grupo de carcinomas neuroendócrinos (NEC) em comparação com outros subtipos. LCC apresentou imunoexpressão proteica inferior aos ADC e SCC, especialmente P53 e VEGF. Detectamos várias "driver" mutações, incluindo EGFR (22%), ERBB2 (2%), genes imunes reguladores CD276 (7%) e CTLA (15%). Observamos também mutações genéticas associadas a à transição epitélio mesenquimal (EMT) como CD44 (31%), MMP2 (3%), VGFA (2%), CDH1 (2.4%), SNAI (3%), VIM (1.2%), e ZEB1 (4%). Curiosamente, uma expressão significativamente maior dos genes AXL e CD44 foi encontrada nos ADCs e SCC em comparação com NEC podendo ser uma evidência de que tais genes possam ser receptores tyrosine kinase (RTK) mais distintos em tais tumores do que nos NEC, sugerindo que esses genes possam ser alvos benéficos no tratamento anti-cancer (73). Mutações no gene BRAF em grandes séries de NSCLC foram investigadas pela primeira vez pelo grupo italiano e foram detectadas em 4.9% dos adenocarcinomas e em 0,3% dos carcinomas de células escamosas (74). A hiperexpressão de HER2 em câncer de pulmão é relatada em cerca de 20% dos casos; todavia a taxa de mutação em uma série de 671 pacientes foi de somente 1,6% em NSCLC, mais frequente em pacientes

que nunca fumaram (3%) e no tipo histológico adenocarcinoma (2,8%) (75). Diversas séries demonstraram frequência variando entre 1% a 6,7% para a fusão ALK-EML4 em populações não selecionadas de NSCLC (76). Inibidores orais específicos de ALK quinase, como por exemplo, o crizotinib, já está em desenvolvimento e excelentes respostas ao seu uso já foram reportadas em pacientes com tumores que apresentam a fusão EML4-ALK: 87% de controle de doença e 57% de resposta objetiva. Os protocolos mais recentes preconizam dois biomarcadores preditivos confirmados para a terapia alvo: mutações do gene EGFR como mais frequente em pacientes que respondem à terapia anti EGFR (Tyrosine Kinase Inhibitors ou TKIs: Gefitinib ou Iressa e Erlotinib ou Tarceva com Cetuximab e Afatinib em confirmação) e a presença de genes de fusão ALK que preveem resposta à terapia anti ALK (Crizotinib) (77).

Contudo, estudos recentes identificaram que a mutação EGFR T790M é responsável pela resistência adquirida aos inibidores de EGFR, em até 50% dos casos e que linhagens de NSCLC que apresentam EGFR do tipo selvagem apresentam sensibilidade aos inibidores de EGFR dependente do grau em que foram submetidos a EMT (15). Em contraste, linhagens de NSCLC resistentes aos inibidores de EGFR com fenótipo mesenquimal apresentaram diminuição da expressão da E-caderina e aumento da expressão da vimentina, sugerindo que EMT pode ser determinante da insensibilidade à inibição do EGFR no câncer de pulmão (70).

XIE e cols (78) demonstraram que o gene NOTCH-1 está envolvido na aquisição de resistência, pelas células tumorais, ao tratamento com drogas inibidoras de tirosina-quinase dos receptores do fator de crescimento epidérmico (EGFR-TKI) em NSCLC. Experimentos em animais demonstram regressão tumoral relacionada à queda da atividade proliferativa e aumento da atividade apoptótica, em resposta a inibição de NOTCH-1 por intermédio da atividade do seu inibidor N-[N-(3,5-difluorophenacetyl)-l-alanyl]-(S)-phenylglycine t-butyl ester (DAPT). Zhang e cols. (79) demonstraram que o mecanismo de ação do TGF-β, através da EMT, pode facilitar a invasão e metástase do adenocarcinoma pulmonar através do evento da EMT por diminuição da expressão gênica de marcadores epiteliais, como a E- caderina, e ao aumento da expressão gênica de marcadores mesenquimais, como a fibronectina e a vimentina. Avaliando a participação de outras moléculas envolvidas na EMT, Merikallio e cols. (80) constataram que houve maior expressão do Snail nos carcinomas de pequenas células em comparação aos carcinomas de células escamosas e aos adenocarcinomas, com importante papel na invasão tumoral e na queda da sobrevida.

Conclui-se que o método de pesquisa (NGS) utilizado neste trabalho permitiu demonstrar que variantes genéticas envolvidas no processo da transição epitélio-mesenquimal e genes dos checkpoints imunológicos estão associados a perfis de células imunes e podem predizer resultados de pacientes e resposta ao bloqueio de checkpoint imunológico. Além disso

demonstrou que as variantes não codificantes de *PD-L1* desempenham um papel importante na modulação da função do checkpoint imunológico e podem ser exploradas como biomarcadores de imunoterapia como a rs7041009, que afeta OS e PPS em pacientes com CPNPC. Finalmente, confirmamos que o NGS fornece informações complementares que orientam o resultado clínico e tratamento de pacientes com CPNPC em estágio inicial portadores de mutações *EGFR* classificadas como mutação de significado clínico desconhecido.

REFERÊNCIAS

1. Bade BC, Dela Cruz CS. Lung Cancer 2020: Epidemiology, Etiology, and Prevention. Clin Chest Med. 2020;41(1):1-24.
2. Instituto Nacional de Câncer (INCA). Tipos de Câncer - Câncer de pulmão [Internet]. 2021 [citado 4 julho 2021]. Disponível em: https://www.inca.gov.br/tipos-de-cancer/cancer-de-pulmao
3. Quintanal-Villalonga A, Paz-Ares L, Ferrer I, Molina-Pinelo S. Tyrosine Kinase Receptor Landscape in Lung Cancer: Therapeutical Implications. Dis Markers. 2016;2016:9214056.
4. Keith RL, Miller YE. Lung cancer chemoprevention: current status and future prospects. Nat Rev Clin Oncol. 2013;10:334-43.
5. Pappu P, Madduru D, Chandrasekharan M, Modhukur V, Nallapeta S, Suravajhala P. Next generation sequencing analysis of lung cancer datasets: A functional genomics perspective. Indian J Cancer. 2016;53:1-7.
6. Veldore VH, Patil S, Satheesh CT, Shashidhara HP, Tejaswi R, Prabhudesai SA, et al. Genomic profiling in a homogeneous molecular subtype of non-small cell lung cancer: An effort to explore new drug targets. Indian J Cancer. 2015;52(2):243-8.
7. Gloeckler Ries LA, Reichman ME, Lewis DR, Hankey BF, Edwards BK. Cancer survival and incidence from the Surveillance, Epidemiology, and End Results (SEER) program. Oncologist. 2003;8(6):541-52.
8. Daniels M, Goh F, Wright CM, Sriram KB, Relan V, Clarke BE, et al. Whole genome sequencing for lung cancer. J Thorac Dis. 2012;4(2):155-63.
9. Pao W, Iafrate AJ, Su Z. Genetically informed lung cancer medicine. J Pathol. 2011;223:230-40.
10. Levy MA, Lovly CM, Pao W. Translating genomic information into clinical medicine: lung cancer as a paradigm. Genome Res. 2012;22:2101-8.
11. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511(7511):543-50.
12. Clinical Lung Cancer Genome Project (CLCGP); Network Genomic Medicine (NGM). A genomics-based classification of human lung tumors. Sci Transl Med. 2013;5(209):209ra153.
13. Bethune G, Bethune D, Ridgway N, Xu Z. Epidermal growth factor receptor (EGFR) in lung cancer: an overview and update. J Thorac Dis. 2010;2:48-51.
14. Sharma SV, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. Nature Rev Cancer. 2007;7:169-81.

15. Xiao D, He J. Epithelial mesenchymal transition and lung cancer. J Thorac Dis. 2010;2:154-9.

16. Yeo MK, Choi SY, Seong IO, Suh KS, Kim JM, Kim KH. Association of PD-L1 expression and PD-L1 gene polymorphism with poor prognosis in lung adenocarcinoma and squamous cell carcinoma. Hum Pathol. 2017;68:103-111.

17. Rose-James A, Tt S. Molecular Markers with Predictive and Prognostic Relevance in Lung Cancer. Lung Cancer Int. 2012;2012:729532.

18. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249.

19. Siegel, RL, Miller, KD, Fuchs, H, Jemal, A. Cancer Statistics, 2021. CA Cancer J Clin. 2021;71:7- 33.

20. Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA). Estimativa 2020: incidência de câncer no Brasil [on-line] Instituto Nacional de Câncer José Alencar Gomes da Silva – Rio de Janeiro: INCA, 2019 [citado 4 julho 2021]. Disponível em: https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//estimativa-2020-incidencia-de-cancer-no-brasil.pdf

21. Corrales L, Rosell R, Cardona AF, Martín C, Zatarain-Barrón ZL, Arrieta O. Lung cancer in never smokers: The role of different risk factors other than tobacco smoking. Crit Rev Oncol Hematol. 2020;148:102895.

22. Dresler CM, León ME, Straif K, Baan R, Secretan B. Reversal of risk upon quitting smoking. Lancet. 2006;368(9533):348–9.

23. Ferreira CG, Rocha JC. *Oncologia Molecular*. 2a ed. 2010. Rio de Janeiro: Atheneu, 2010.

24. Akhtar N, Bansal JG. Risk factors of Lung Cancer in nonsmoker. Curr Probl Cancer. 2017;41(5):328-39.

25. Smith CJ, Perfetti TA, Rumple MA, Rodgman A, Doolittle DJ. "IARC group 2A Carcinogens" reported in cigarette mainstream smoke. Food Chem Toxicol. 2000;38(4):371-83.

26. Smith CJ, Perfetti TA, Rumple MA, Rodgman A, Doolittle DJ. “IARC group 2B Carcinogens” reported in cigarette mainstream smoke. Food Chem Toxicol. 2001;39(2):183-205.

27. Herbst RS, Heymach JV, Lippman SM. Lung cancer. N Engl J Med. 2008;359(13):1367-80.

28. Pikor LA, Ramnarine VR, Lam S, Lam WL. Genetic alterations defining NSCLC subtypes and their therapeutic implications. Lung Cancer. 2013;82(2):179-189

29. Gazdar AF, Bunn PA, Minna JD. Small-cell lung cancer: What we know, what we need to know and the path forward. Nature Reviews Cancer. 2017;17(12):725-737

30. Novaes FT, Cataneo DC, Ruiz Junior RL, Defaveri J, Michelin OC, Cataneo AJM. Câncer de pulmão: histologia, estádio, tratamento e sobrevida. J Bras Pneumol. 2008;34(8):595-600.

31. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: A heterogeneous set of diseases. Nature Rev Cancer. 2014;14(8):535-546.

32. Barros-Filho MC, Guisier F, Rock LD, Becker-Santos DD, Sage AP, Marshall EA, et al. Tumour Suppressor Genes with Oncogenic Roles in Lung Cancer [Internet]. In: Lemamy G-J. *Genes and Cancer*. IntechOpen; 2019. Disponível em: https://www.intechopen.com/chapters/65960

33. Muller, K. Histological classification and histogenesis of lung-cancer. Eur J Respir Dis. 1984;65:4–19.

34. Andrade FM, Mourad OMA, Judice LF. Revisão do sistema de estadiamento tumor-nódulo-metástase para câncer de pulmão: mudanças e perspectivas. J Bras Pneumol. 2010;36(5): 617-620.

35. Reck M, Rabe KF. Precision Diagnosis and Treatment for Advanced Non-Small-Cell Lung Cancer. N Engl J Med. 2017;377(9):849-861.

36. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. J Thorac Oncol. 2016;11(1):39-51.

37. Araujo LH, Baldotto C, Castro Jr G, Katz A, Ferreira CG, Mathias C, et al. Lung cancer in Brazil. J Bras Pneumol. 2018;44(1):55-64.

38. Sá VK. Biomarcadores de prognóstico no câncer de pulmão: caracterização do perfil de expressão genica das hialuronidades, imunoreatividade das das hialuronidades e sintases do ácido hialurônico e interação dessas proteínas com a transição epitélio-mesenquimal [tese]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2012.

39. Baldotto C, Masson P, Zukin M, Araujo LH. Mutações drivers em câncer de pulmão não-pequenas células (CPNPC). Pulmão RJ. 2016;25(2):23-28.

40. Sequist LV, Heist RS, Shaw AT, Fidias P, Rosovsky R, Temel JS, et al. Implementing multiplexed genotyping of non-small-cell lung cancers into routine clinical practice. Ann Oncol. 2011;22(12):2616-2624.

41. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. Lancet Oncol. 2011;12(2):175-80.

42. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511(7511):543-50.

43. Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, Maiello MR, et al. Epidermal growth factor receptor (egfr) signaling in cancer. Gene. 2006;366:2–16.

44. Sibilia M, Kroismayr R, Lichtenberger BM, Natarajan A, Hecking M, Holcmann M. The epidermal growth factor receptor: from development to tumorigenesis. Differentiation. 2007;75(9):770–787

45. Red Brewer M, Choi SH, Alvarado D, Moravcevic K, Pozzi A, Lemmon MA, et al. The juxtamembrane region of the EGF receptor functions as an activation domain. Mol Cell. 2009;34:641–651.

46. Gazdar AF. Epidermal growth factor receptor inhibition in lung cancer: the evolving role of individualized therapy. Cancer Metastasis Rev. 2010;29(1):37-48.

47. Linggi B, Carpenter G. Erbb receptors: new insights on mechanisms and biology. Trends Cell Biol. 2006;16(12):649-656.

48. Adamo V, Franchina T, Adamo B, Denaro N, Gambadauro P, Chiofalo G, et al. Gefitinib in lung cancer therapy: clinical results, predictive markers of response and future perspectives. Cancer Biol Ther. 2009;8(3):206-12.

49. Zhang Z, Stiegler AL, Boggon TJ, Kobayashi S, Halmos B. EGFR-mutated lung cancer: a paradigm of molecular oncology. Oncotarget. 2010;1:497-514.

50. Schiller JH, Harrington D, Belani C, Langer C, Sandler A, Krook J, et al. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. N Engl J Med. 202;346: 92–98.

51. Lopes GL, Vattimo EF, Castro Junior GD. Identifying activating mutations in the EGFR gene: prognostic and therapeutic implications in non-small cell lung cancer. J Bras Pneumol. 2015;41(4):365-375.

52. Brasil. Presidência da República. Secretaria da Saúde. Portaria n. 52 de 7 de novembro de 2013. Diário Oficial da União, Brasília (DF). 2013. [citado 8 julho 2021] Disponível em: https://bvsms.saude.gov.br/bvs/saudelegis/sctie/2013/prt0052_07_11_2013.html

53. Araujo LH, Castro Jr G, Baldotto C, Custódio MG, De-Marchi P, Calabrich A, et al. Efficacy of osimertinib in advanced T790M-positive NSCLC after progression to prior EGFR-TKI: real world data from a Brazilian cohort. Braz J Oncol. 2020;16:e-20200005

54. ANVISA. Nova indicação para Câncer de Pulmão (CPNPC) [Internet]. 2018 [citado 9 julho 2021]. Disponível em: https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2018/nova-indicacao-para-cancer-de-pulmao-cpnpc

55. Garassino MC, Martelli O, Broggini M, Farina G, Veronese S, Rulli E, et al. Erlotinib versus docetaxel as second-line treatment of patients with advanced non-small-cell lung cancer and wild-type EGFR tumours (TAILOR): a randomised controlled trial. Lancet Oncol. 2013;14(10):981-8.
56. Kalluri R, Weinberg RA. The basics of epitelial-mesenchymal transition. J Clin Invest. 2009;119(6):1420–1428.
57. Acloque H, Adams MS, Fishwick K, Bronner-Fraser M, Nieto MA. Epithelial-mesenchymal transitions: the importance of changing cell state in development and disease. J Clin Invest. 2009;119(6):1438-49.
58. Georgakopoulos-Soares I, Chartoumpekis DV, Kyriazopoulou V, Zaravinos A. EMT Factors and Metabolic Pathways in Cancer. Front Oncol. 2020;10:499.
59. Creighton CJ, Gibbons DL, Kurie JM. The role of epithelial-mesenchymal transition programming in invasion and metastasis: a clinical perspective. Cancer Manag Res. 2013;5:187-95.
60. Weber CE, Li N Y, Wai PY, Kuo PC. Epithelial-mesenchymal transition, TGF-β, and osteopontin in wound healing and tissue remodeling after injury. J Burn Care Res. 2012;33(3):311–318.
61. López-Novoa JM, Nieto MA. Inflammation and EMT: an alliance towards organ fibrosis and cancer progression. EMBO Mol Med. 2009;1(6-7):303-14.
62. Cursons J, Pillman KA, Scheer KG, Gregory PA, Foroutan M, Hediyeh-Zadeh S, et al. Combinatorial targeting by microRNAs co-ordinates post-transcriptional control of EMT. Cell Syst. 2018;7:77–91.e7.
63. Korpal M, Lee ES, Hu G, Kang Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. J Biol Chem. 2008;283:14910–4.
64. Beers MF, Morrisey EE. The three R's of lung health and disease: repair, remodeling, and regeneration. The Journal of Clinical Investigation. 2011;121(6):2065-73.
65. Capelozzi VL. Oncologia Torácica: Classificação Histopatológica. Medicina Respiratória, Eds. Carlos Alberto de Castro Pereira e Marcelo Alcantara Holanda. Editora Atheneu 2014;Vol 2:pp 1037.
66. Allavena P, Sica A, Solinas G, Porta C, Mantovani A. The inflammatory micro-environment in tumor progression: The role of tumor-associated macrophages. Critical Reviews in Oncology/Hematology. 2008;66(1):1-9.
67. Akira S, Takeda K. Toll-like receptor signalling. Nat Rev Immunol 2004;4:499 –511.

68. Walzer T, Jaeger S, Chaix J, Vivier E. Natural killer cells: from CD3−NKp46+ to post-genomics meta-analyses. Current Opinion in Immunology. 2007;19(3):365-72.

69. Kalluri R, Neilson EG. Epithelial-mesenchymal transition and its implications for fibrosis. The Journal of Clinical Investigation. 2003;112(12):1776-84.

70. Byers LA, Diao L, Wang J, Saintigny P, Girard L, Peyton M, et al. An Epithelial–Mesenchymal Transition Gene Signature Predicts Resistance to EGFR and PI3K Inhibitors and Identifies Axl as a Therapeutic Target for Overcoming EGFR Inhibitor Resistance. Clinical Cancer Research. 2013;19(1):279-90.

71. Wu Y, Zhou BP. New insights of epithelial-mesenchymal transition in cancer metastasis. Acta Biochim Biophys Sin (Shanghai). 2008;40(7):643-50.

72. Nagelkerke A, Bussink J, Rowan AE, Span PN. The mechanical microenvironment in cancer: How physics affects tumours. Seminars in Cancer Biology. 2015;35:62-70.

73. Machado J, Ascheri D, Leao P, et al.Protein and molecular alterations in EMT pathways of lung cancer: a comparative analysis between NSCLC. Journal of Thoracic Oncology 2017; 12 (S1): S420.

74. Marchetti A, Felicioni L, Malatesta S, Grazia Sciarrotta M, Guetti L, Chella A, et al. Clinical Features and Outcome of Patients With Non–Small-Cell Lung Cancer Harboring BRAF Mutations. Journal of Clinical Oncology. 2011;29(26):3574-9.

75. Shigematsu H, Takahashi T, Nomura M, Majmudar K, Suzuki M, Lee H, et al. Somatic Mutations of the HER2 Kinase Domain in Lung Adenocarcinomas. Cancer Research. 2005;65(5):1642-6.

76. Bang Y-J. Treatment of ALK-Positive Non–Small Cell Lung Cancer. Archives of Pathology & Laboratory Medicine. 2012;136(10):1201-4.

77. Rodriguez-Canales J, Parra-Cuentas E, Wistuba II. Diagnosis and Molecular Classification of Lung Cancer. Cancer Treat Res. 2016;170:25-46.

78. Xie M, He C-S, Wei S-H, Zhang L. Notch-1 contributes to epidermal growth factor receptor tyrosine kinase inhibitor acquired resistance in non-small cell lung cancer *in vitro* and *in vivo*. European Journal of Cancer. 2013;49(16):3559-72.

79. Zhang HJ, Wang HY, Zhang HT, et al. Transforming growth factor-β1 promotes lung adenocarcinoma invasion and metastasis by epithelial-to-mesenchymal transition. Mol Cell Biochem 2011; 1-2: 309-314

80. Merikallio H, Turpeenniemi-Hujanen T, Pääkkö P, Mäkitaro R, Riitta K, Salo S, et al. Snail promotes an invasive phenotype in lung carcinoma. Respir Res. 2012;13(1):104.

APÊNDICE

Materiais e Métodos

Casuística

A casuística foi composta por 70 amostras de tecido tumoral primário *congelado* de CPNPC, provenientes do Hospital da Clínicas de São Paulo (HCFMUSP), do Instituto do Câncer de São Paulo (ICESP), do Instituto do Coração (Incor) e do Biobanco do A.C. Camargo Cancer Center sob condições de armazenamento de rotina. Os pacientes foram submetidos à ressecção cirúrgica entre janeiro de 1995 e dezembro de 2015. Os critérios de exclusão da casuística foram pacientes com uma segunda neoplasia, doenças genéticas ou síndromes genéticas prévias. Os dados demográficos, parâmetros clínicos, informações sobre a terapia e sobrevida global foram recuperados dos registros médicos. A sobrevida global foi definida como o período compreendido entre a data da cirurgia até a data do último acompanhamento ou óbito. O estudo foi aprovado de acordo com as normas do Comitê de Ética em Pesquisa da Faculdade de Medicina da Universidade de São Paulo (CAPPEsq) - CAAE: 79769017.1.0000.5440; Parecer Consubstanciado Número 2.673.320/2018). A dispensa do Termo de Consentimento Livre e Esclarecido (TCLE) foi obtida do comitê, e a identidade dos participantes nesta análise retrospectiva foi omitida e anonimizada.

Procedimentos laboratoriais

Extração de DNA

As amostras foram submetidas à extração de DNA genômico (DNAg) utilizando o kit comercial QIAamp DNA mini® Kit (Qiagen) seguindo as recomendações do fabricante.

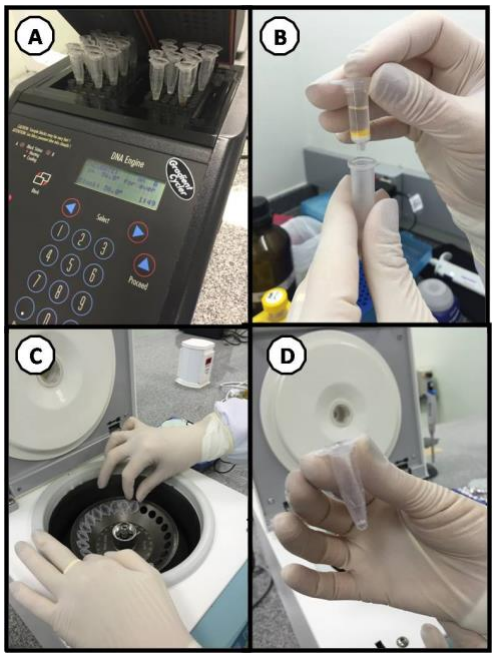


Figura 1: Extração de DNA pelo kit da Qiagen. A – Aquecimento. B – Ligação do DNA na membrana. C – Lavagem e centrifugação. D – Eluição. Fonte: Experimentos do Projeto.

Quantificação de DNA

Após a extração, as amostras de DNAg foram analisadas em relação a sua concentração e pureza determinadas através da leitura em espectrofotômetro (NanoDrop® 2000) indicando pelo espectro de absorbância contaminação por proteína (260/280) e/ou por fenol e minerais (260/230). Posteriormente, o DNA foi quantificado por fluorescência em equipamento Qubit® 2.0 Fluorometer (Invitrogen™) utilizando o Qubit dsDNA HS Assay kit (Life Technologies). Todas as amostras foram normalizadas para 25 ng/mL.

Construção do painel genético

O painel foi desenvolvido pela ferramenta online Design Studio (Illumina) baseado na tecnologia TruSeq® Custom Amplicon v1.5 (TSCA) (Illumina). Com a ajuda do suporte técnico-científico da Illumina, o painel foi customizado com 764 amplicons de 250 pares de base (pb) para investigar as regiões: região promotora, regiões codificantes, introns, 5'UTR e 3'UTR de 38 genes: *ANKFN1, AXL, BAI1, BRAF, CD274, CD276, CD44, CDH1, CTLA-4, EGFR, EPHA5, ERBB2, KCNIP4, KLLN, KRAS, LAG3, MAP2K1, MMP2, MUC1, NRAS, PDCD1LG2, PIK3CA, PTEN, PTENP1, RAB25, SNAI1, SNAI2, SNAPC5, SND1, TGFB1, TP53, VEGFA, VIM, VTCN1, WRAP53, ZEB1, ZEB2, ZEB2-A1*. O painel forneceu uma cobertura total de 96% em toda a região de interesse.

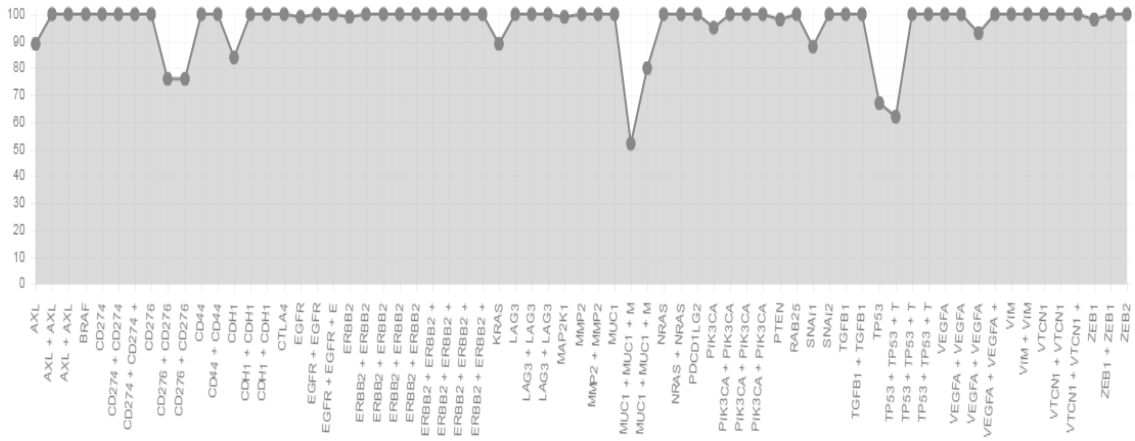


Figura 2: Gráfico do Índice de Cobertura. Índice de cobertura do painel TruSeq® Custom Amplicon v1.5 das regiões-alvo. Fonte: TSCA Design Detail Report Illumina.

Construção das bibliotecas

Para a construção das bibliotecas de DNA foi empregado o protocolo TruSeq® Custom Amplicon v1.5 Reference Guide (Documento #15027983 v02), permitindo sequenciar até 96 amostras em uma reação multiplex.

Hibridização dos oligonucleotídeos

Para estudar as regiões gênicas contidas no painel, conjuntos de oligonucleotídeos específicos foram adicionados ao DNAg para a hibridização na região de interesse. A reação foi incubada a 37°C por 1 minuto seguida por 40°C a 80 minutos.

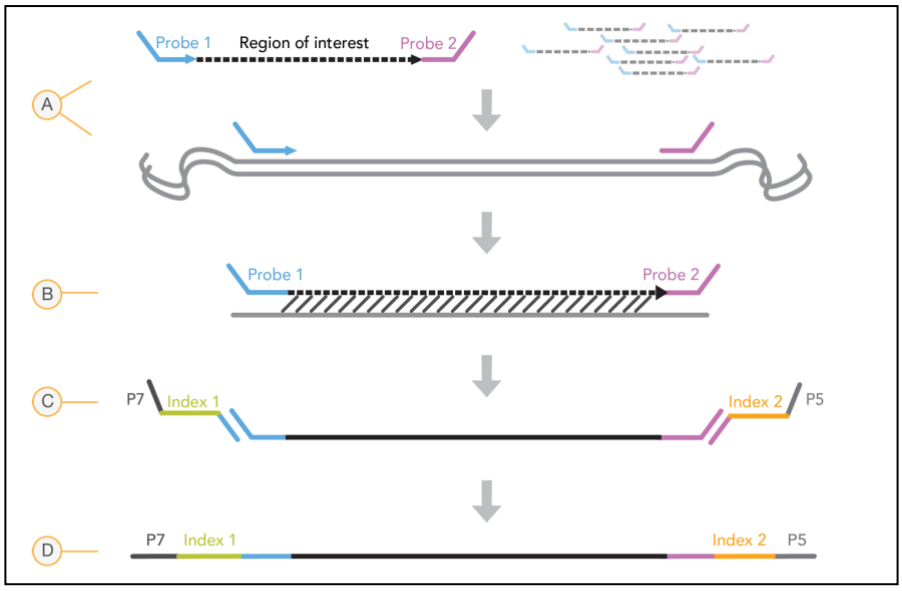


Figura 3: Esquema da preparação da biblioteca de DNA para NGS. A- Hibridação dos oligonucleotídeos; B- Extensão e ligação; C- Adição dos índices e adaptadores; D- amplicon pronto para o sequenciamento. Fonte: www.illumina.com

Remoção dos oligonucleotídeos não ligados

Após a hibridização, a reação foi transferida para placa-filtro e tampões de lavagem foram adicionados para a remoção do excesso de oligonucleotídeos.

Extensão e ligação dos oligonucleotídeos ligados

Após a remoção de oligonucleotídeos não ligados, o DNA hibridizado foi incubado a 37°C por 45 minutos em um tampão contendo DNA polimerase, responsável pela extensão da região-alvo a partir do oligonucleotídeo à extremidade da região 3' (upstream), e DNA ligase, responsável pela ligação à extremidade 5' do oligo (downstream), formando os produtos contendo as regiões de interesse necessários para a amplificação.

Amplificação das bibliotecas

Os produtos contendo as regiões de interesse foram amplificados por reação em cadeia da polimerase (PCR). Para a reação foram utilizados reagentes contendo sequências índices (i5 e i7) para a multiplexação das amostras e adaptadores (P5 e P7) necessários para a formação de clusters no sequenciamento. A condição de ciclagem da reação foi de: 3 minutos a 95°C, seguidos de 24 ciclos de 30 segundos a 95°C, 30 segundos a 66°C, 60 segundos a 72°C, com extensão final de 5 minutos a 72°C.

Purificação das bibliotecas

Após a reação de PCR, as bibliotecas foram purificadas com lavagens de etanol a 80% e uso de esferas magnéticas Agencourt® AMPure® XP beads para retirar o excesso de sequências índices e adaptadores.

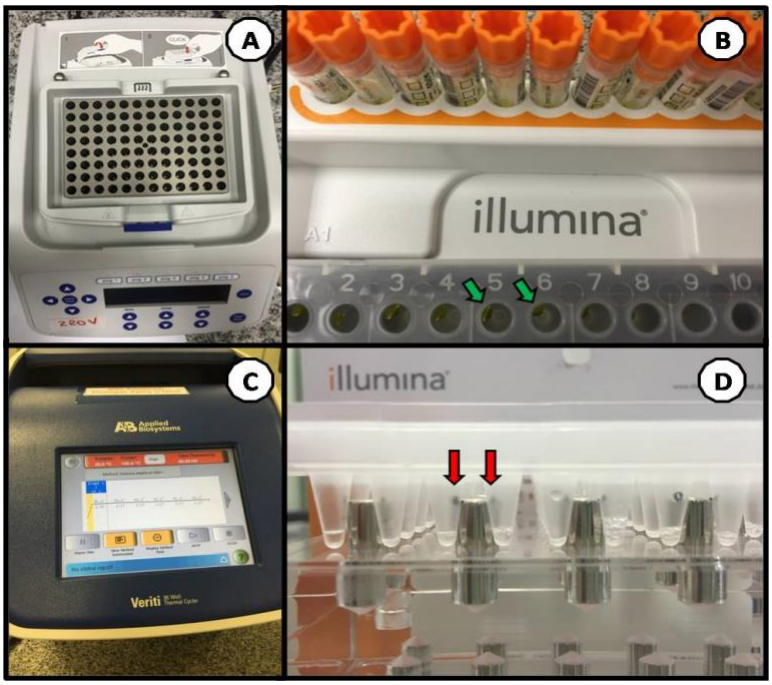


Figura 4: Construção da biblioteca de DNA. A – Termobloco para hibridização dos oligos. B – Indexação da biblioteca. Note o conteúdo colorido de cada índex (setas verdes). C – Amplificação da biblioteca por PCR. D – Clean-up com beads magnéticos. Observa-se o DNA condensado junto à coluna metálica (setas vermelhas). Fonte: nossos experimentos.

Validação e Normalização das bibliotecas

Seguida a purificação, a técnica de eletroforese em gel de poliacrilamida a 7% confirmou a amplificação de fragmentos de aproximadamente 350 pb, sendo 250pb da biblioteca de amplicon mais 100 pb de índices e adaptadores. Em seguida, as bibliotecas foram normalizadas a concentração de 4nM conforme protocolo do fabricante. As bibliotecas normalizadas foram combinadas em volumes iguais em um mesmo tubo formando um pool, o qual foi diluído a 1,8pM e desnaturado para a execução do sequenciamento.

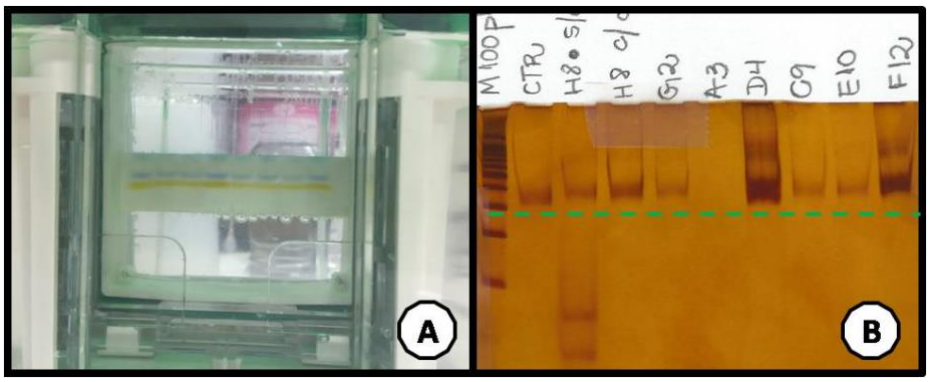


Figura 5: Validação da biblioteca com amplicons de 350pb. Gel de poliacrilamida 7% (Marcador molecular 100pb). A – Eletroforese em corrida. B – Bandas no gel indicando 350 pb em cada amostra (linha tracejada em verde). Fonte: nossos experimentos.

Sequenciamento

O pool de bibliotecas foi submetido ao sequenciamento na plataforma Miseq® Illumina configurada para uma corrida de 300 ciclos do tipo paired-end (2 x 150) (Figura 6). A duração da execução levou aproximadamente 24 horas.

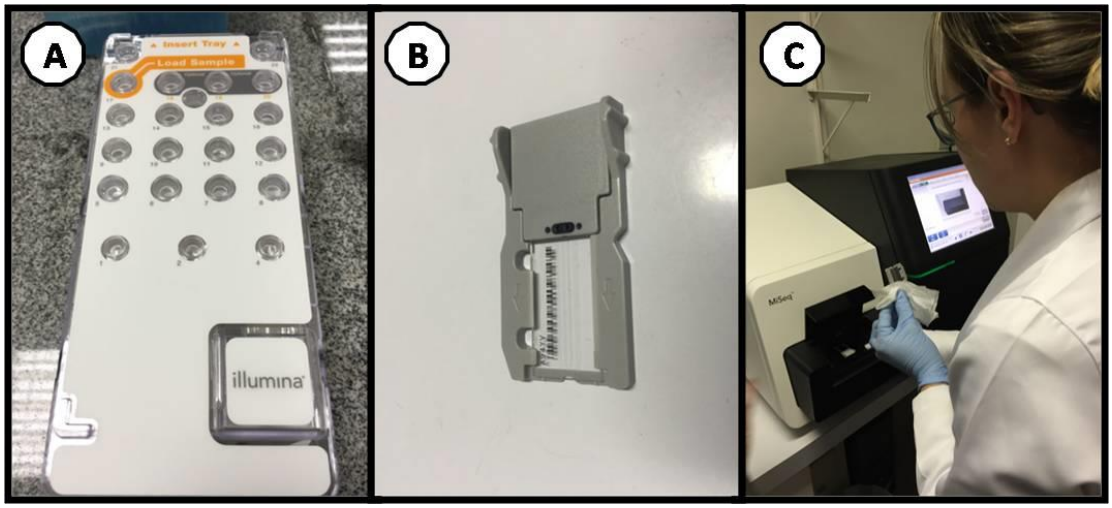


Figura 6: Corrida de Sequenciamento. A – Cartucho com reagentes e biblioteca para sequenciamento. B – Flow cell. C – Carregando o MiSeq. Fonte: nossos experimentos.

Análise de bioinformática

A análise dos dados de sequenciamento brutos foi realizada no próprio sistema MiSeq® utilizando o software MiSeq Reporter v.1.8.1 (Illumina, San Diego, CA, EUA) com parâmetros padrão. As leituras filtradas foram alinhadas ao genoma humano (hg19, GRCh37) usando a ferramenta Burrows-Wheeler Alignment (BWA) v.0.7.10. Após o alinhamento, o software SAMtools foi aplicado para converter os arquivos de alinhamento em um formato de mapa de alinhamento binário indexado (BAM). As variantes de nucleotídeo único (SNVs) e as pequenas inserções e deleções (INDELs) foram chamadas usando o GATK UnifiedGenotyper incluindo HaplotypeCaller com parâmetros padrão baseados em hg19 e anotados com dbSNP versão 144 (https://www.ncbi.nlm.nih.gov/ SNP /). As variantes foram anotadas usando o software VEP com base nas consequências, impactos previstos e frequências alélicas relatadas na população. As variantes foram identificadas com base nos seguintes critérios de corte: número de leituras com a base alterada no tumor ≥10; variantes detectadas em uma posição de profundidade total de leitura ≥100 ×; variantes com uma frequência populacional superior a 1% (popfreq_all> 0,01) disponíveis publicamente nos bancos de dados foram consideradas polimorfismos e inferior a 1% foram consideradas mutações somáticas.

ANEXOS

Variants in Epithelial-Mesenchymal Transition and Immune Checkpoint Genes Are Associated With Immune Cell Profiles and Predict Survival in Non–Small Cell Lung Cancer

Edwin Roger Parra, MD, PhD; Mei Jiang, PhD; Juliana Machado-Rugolo, PhD; Lygia Bertalha Yaegashi, MD, PhD; Tabatha Prieto, PhD; Cecília Farhat, PhD; Vanessa Karen de Sá, PhD; Maria Aparecida Nagai, PhD; Vladmir Cláudio Cordeiro de Lima, MD; Tereza Takagaki, MD, PhD; Ricardo Terra, MD, PhD; Alexandre Todorovic Fabro, MD, PhD; Vera Luiza Capelozzi, MD, PhD

• ***Context.***—**Identification of gene mutations that are indicative of epithelial-mesenchymal transition and a noninflammatory immune phenotype may be important for predicting response to immune checkpoint inhibitors.**

Objective.—**To evaluate the utility of multiplex immunofluorescence for immune profiling and to determine the relationships among tumor immune checkpoint and epithelial-mesenchymal transition genomic profiles and the clinical outcomes of patients with nonmetastatic non–small cell lung cancer.**

Design.—**Tissue microarrays containing 164 primary tumor specimens from patients with stages I to IIIA non–small cell lung carcinoma were examined by multiplex immunofluorescence and image analysis to determine the expression of programmed death ligand-1 (PD-L1) on malignant cells, $CD68^+$ macrophages, and cells expressing the immune markers CD3, CD8, CD57, CD45RO, FOXP3, PD-1, and CD20. Immune phenotype data were tested for correlations with clinicopathologic characteristics, somatic and germline genetic variants, and outcome.**

Results.—**A high percentage of $PD\text{-}L1^+$ malignant cells was associated with clinicopathologic characteristics, and high density of $CD3^+PD\text{-}1^+$ T cells was associated with metastasis, suggesting that these phenotypes may be clinically useful to identify patients who will likely benefit from immunotherapy. We also found that *ZEB2* mutations were a proxy for immunologic ignorance and immune tolerance microenvironments and may predict response to checkpoint inhibitors. A multivariate Cox regression model predicted a lower risk of death for patients with a high density of $CD3^+CD45RO^+$ memory T cells, carriers of allele G of *CTLA4* variant rs231775, and those whose tumors do not have *ZEB2* mutations.**

Conclusions.—**Genetic variants in epithelial-mesenchymal transition and immune checkpoint genes are associated with immune cell profiles and may predict patient outcomes and response to immune checkpoint blockade.**

(*Arch Pathol Lab Med.* 2020;144:1234–1244; doi: 10.5858/arpa.2019-0419-OA)

Accepted for publication January 8, 2020.

Published online March 9, 2020.

Supplemental digital content is available for this article at www.archivesofpathology.org in the October 2020 table of contents.

From the Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston (Parra, Jiang); the Department of Pathology and Laboratory of Genomics and Histomorphometry (Machado-Rugolo, Yaegashi, Prieto, Farhat, de Sá, Capelozzi) and the Division of Pneumology, Heart Institute (Incor) (Takagaki), Faculty of Medicine, University of São Paulo, São Paulo, Brazil; the Department of Oncology, Clinicas Hospital, Faculty of Medicine, São Paulo State University, São Paulo, Brazil (de Sá, Nagai); the Medical Oncology Department and Translational Immune-Oncology Group, A. C. Camargo Cancer Center, Sã Paulo, Brazil (de Sá, de Lima); the Department of Thoracic Surgery, Institute of Cancer of São Paulo, São Paulo, Brazil (Terra); the Department of Thoracic Surgery, Heart Institute (Incor), São Paulo, Brazil (Terra); and the Department of Pathology and Legal Medicine, Ribeirão Preto School of Medicine, University of São Paulo, Ribeirão Preto, Brazil (Fabro).

Supported by The University of Texas Lung Cancer Specialized Programs of Research Excellence (grant P50CA70907) and the Foundation for the Support of Research of the State of São Paulo (FAPESP 2013/14277-4, FAPESP 2018-20403-6).

The authors have no relevant financial interest in the products or companies described in this article.

Corresponding author: Edwin Roger Parra, MD, PhD, Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Unit 951, Houston, TX 77030 (email: erparra@mdanderson.org).

Non–small cell lung cancer (NSCLC) accounts for about 85% of all lung cancers. The prognosis of patients with NSCLC is highly dependent on the tumor-node-metastasis (TNM) stage; as the volume of disease increases, survival declines dramatically.[1] Surgery is the standard treatment for patients with stage I and stage II NSCLC and for selected patients with stage IIIA disease with the possibility of complete tumor resection. The addition of adjuvant cisplatin-based chemotherapy to surgery improved 5-year survival rates by 5% to 10%, but no significant therapeutic innovation has been established thereafter, and the overall 5-year survival rate remains below 50%.

Over the past few years, immunotherapy with immune checkpoint inhibitors has led to remarkable progress in non–oncogene-addicted metastatic NSCLC.[2] Most of these new immunotherapeutic approaches for advanced NSCLC are now being studied in the context of nonmetastatic

NSCLC.[3] However, not all NSCLC patients receive the same benefit from immunotherapy as a first or second line of treatment. To date, programmed death ligand-1 (PD-L1) is the only validated biomarker for immunotherapy response in NSCLC patients. However, this biomarker is not sufficiently robust and poses many challenges. For example, some patients with more than 50% PD-L1$^+$ tumor cells do not have a response to anti–programmed death receptor-1 (PD-1)/PD-L1 treatment; conversely, some patients whose tumors do not express PD-L1 see good responses.[4–6]

Some studies have suggested that the heterogeneity of PD-1/PD-L1 axis expression and response to PD-1/PD-L1 treatment in NSCLC depends on alterations in other factors, such as immune evasion mechanisms and the tumor immune microenvironment. Expression of genes involved in the epithelial-mesenchymal transition (EMT) and immune checkpoints are also known to control cancer progression.[7–16] However, the associations among these factors and their association with immunotherapy response remain to be determined.

We hypothesized that genetic variants that are indicative of EMT and host immune factors predict patients' outcomes and responses to immune checkpoint blockade. Identification of mutations that can serve as a proxy for a noninflammatory immune phenotype would further enable us to predict response to immune checkpoint blockade. Such information would be particularly important in clinical settings where next-generation sequencing (NGS) is available, but comprehensive studies of tumor-associated immune cells (TAICs) are not. Bearing this in mind, we determined the extent to which immune and genomic profiles predicted clinical outcomes of patients with NSCLC.

PATIENTS AND METHODS

Cases and Specimens

We obtained archival formalin-fixed, paraffin-embedded histologic tumor sections from 164 patients diagnosed with NSCLC who had undergone surgical resection between January 1, 1995, and December 31, 2015, and who had not received neoadjuvant therapy. Patients had been treated at the Hospital das Clinicas of the State University of São Paulo Medical School, the Heart Institute of the University of São Paulo, the Cancer Institute of São Paulo, and the A.C. Camargo Cancer Center, São Paulo, Brazil. The histologic diagnoses were reviewed by 2 experienced lung pathologists, who judged their accuracy based on the World Health Organization 2015 classification system.[17] The specimens included 94 adenocarcinomas, 51 squamous cell carcinomas (SCCs), and 19 large cell carcinomas (LCCs). Tissue microarrays were constructed from the primary resected tumors: 3 cores (diameter, 1.5 mm) of representative tumor tissue were taken from a representative formalin-fixed, paraffin-embedded block from each patient case. Clinicopathologic data, including pathologic TNM staging, which was determined according to the guidelines of the International Association for the Study of Lung Cancer, 8th edition,[1] were obtained from medical records (Table 1).

Our study received approval from the institutional review boards of the institutions involved in the project (process number H-1404-100-572). A waiver of the requirement for informed consent was obtained from the review boards of all participating institutions.

Multiplex Immunofluorescence Staining

Multiplex immunofluorescence (mIF) staining was performed using methods similar to previously described and validated ones.[18] Consecutive 4-μm-thick tissue microarray sections were stained using an automated staining system (BOND-RX, Leica Biosystems, Buffalo Grove, Illinois) using 2 different panels. Panel 1 consisted of pancytokeratin AE1/AE3 (dilution 1:300; epithelial cell positive, Dako, Carpinteria, California), PD-L1 (dilution 1:100; clone E1L3N, Cell Signaling Technology, Danvers, Massachusetts), PD-1 (dilution 1:250; clone EPR4877-2, Abcam, Cambridge, Massachusetts), CD3 (dilution 1:100; T-cell lymphocytes, Dako), CD8 (dilution 1:20; cytotoxic T cells, clone C8/144B, Thermo Fisher Scientific, Waltham, Massachusetts), and CD68 (dilution 1:450; macrophages, clone PG-M1, Dako). Panel 2 consisted of AE1/AE3 (epithelial cell-positive, dilution 1:300; Dako), CD3 (dilution 1:100; T cell lymphocytes, Dako), CD57 (dilution 1:40; natural killer cells, clone HNK-1, BD Biosciences, San Jose, California), CD45RO (ready to use; memory T cells, clone UCHL1, Leica Biosystems), FOXP3 (dilution 1:50; regulatory T cells, clone 206D, BioLegend, San Diego, California), and CD20 (dilution 1:100; B-cell lymphocytes, Dako). All the markers were stained in sequence using their respective fluorophores, which were contained in the Opal 7 kit (Cat #NEL797001KT, Akoya Biosciences/PerkinElmer, Waltham, Massachusetts) after IF validation to obtain a uniform, specific, and correct signal across all channels and ensure a well-balanced staining pattern during the multiplex staining.[18] The correct signal from each fluorophore was also defined and optimized between 10 and 30 counts of intensity to maintain a good balance and similar thresholds of intensity across all the antibodies. In parallel, to detect possible variations in staining and optimal separation of the signal, positive and negative (autofluorescence) controls were included during the staining. Autofluorescence controls with an expected spectrum of 488 nm can accurately remove the autofluorescence from all the labeling signals during the analysis process. The stained slides were scanned using a multispectral microscope (Vectra Polaris 3.0 imaging system, Akoya Biosciences/PerkinElmer) under fluorescence conditions (Supplemental Figure 1; see supplemental digital content containing 4 tables and 1 figure at www.archivesofpathology.org in the October 2020 table of contents).

mIF Quantitation

Multispectral images of tumor sections from each core were analyzed with inForm 2.2.1 software (Akoya Biosciences/PerkinElmer). Individual cells, which were defined by nuclear staining and identified by the inForm cell segmentation tool, were used in a phenotype pattern-recognition learning algorithm to characterize the colocalization of the cell populations labeled with 2 panels.[19] Panel 1 labeling was as follows: (1) malignant cells (MCs) with the AE1/AE3$^+$ marker, including those with and without PD-L1 expression (AE1/AE3$^+$PD-L1$^+$ and AE1/AE3$^+$PD-L1$^-$, respectively); (2) T lymphocytes, including CD3$^+$ and pan–T-cell markers (cytotoxic T cells [CD3$^+$CD8$^+$], antigen-experienced T cells [CD3$^+$PD-1$^+$], and other CD3$^+$ T cells); (3) CD68$^+$ macrophages; and (4) macrophages expressing PD-L1 (CD68$^+$PD-L1$^+$). Panel 2 labeling was as follows: (1) MCs with the AE1/AE3$^+$ marker; (2) T lymphocytes, including CD3$^+$ and pan–T-cell markers (memory T cells [CD3$^+$CD45RO$^+$], natural killer T cells [CD3$^+$CD57$^+$], memory/regulatory T cells

Table 1. Clinicopathologic Characteristics of 164 Patients With Non–Small Cell Lung Cancer			
Characteristic	Adenocarcinoma (n = 94)	SCC (n = 51)	LCC (n = 19)
Age, median, y	66	67	59
Sex, No. (%)			
Female	46 (49)	19 (37)	6 (32)
Male	48 (51)	32 (63)	13 (68)
Tobacco history, No. (%)			
No	19 (20)	1 (2)	1 (5)
Yes	51 (54)	48 (94)	7 (37)
Unknown	24 (26)	2 (4)	11 (58)
Tumor size, median, cm	3.0	3.2	5.0
Tumor category, No. (%)[a]			
T1	29 (30)	14 (27)	6 (32)
T2	47 (51)	30 (59)	13 (68)
T3	15 (16)	5 (10)	0 (0)
T4	3 (3)	2 (4)	0 (0)
Nodal status, No. (%)[a]			
N0	82 (87)	41 (80)	10 (53)
N1	12 (13)	10 (20)	9 (47)
IASLC pathologic stage, No. (%)[a]			
I	44 (47)	30 (59)	10 (53)
II	29 (31)	11 (21)	5 (26)
IIIA	21 (22)	10 (20)	4 (21)
Disease progression, No. (%)	16 (17)	12 (24)	2 (11)
Adjuvant treatment, No. (%)			
No	74 (79)	41 (80)	15 (79)
Yes	20 (21)	10 (20)	4 (21)
Patients censored for survival analysis at last time of follow-up, No. (%)	32 (34)	15 (29)	9 (47)

Abbreviations: IASLC, International Association for the Study of Lung Cancer; LCC, large cell carcinoma; SCC, squamous cell carcinoma.
[a] IASLC staging guidelines, 8th edition.[1]

[$CD3^+CD45RO^+FOXP3^+$], and other $CD3^+$ T cells); and (3) B-cell lymphocytes ($CD20^+$).

The individual cell phenotype report produced by the inForm software was processed using Microsoft Excel 2010, and a final summary of the data, which contained the median number of cells/mm^2 for each phenotype and the median percentage of macrophages and MCs expressing PD-L1, was created for statistical analysis. If the percentage of MCs or macrophages expressing PD-L1 was greater than the median value, PD-L1 expression was considered positive. If the percentage of macrophages or MCs expressing PD-L1 in a sample was lower than or equal to the median, PD-L1 expression was considered negative. Likewise, if the density of TAICs in a cell population was greater than the median, it was considered high density; if the density was less than or equal to the median, it was considered low.

In addition, as proposed by Teng and colleagues,[20] we identified 4 different types of tumor microenvironments on the basis of the density of $CD3^+$ tumor-infiltrating lymphocytes (TILs) (a density greater than the median was considered high) and the PD-L1 expression level of MCs. The 4 microenvironment types were type I, adaptive immune resistance ($PD\text{-}L1^+/TIL^+$); type II, immunologic ignorance ($PD\text{-}L1^-/TIL^-$); type III, intrinsic induction ($PD\text{-}L1^+/TIL^-$); and type IV, tolerance ($PD\text{-}L1^-/TIL^+$).

Genomic Analysis

Following previously published procedures,[21] we extracted DNA from fresh-frozen lung cancer tissue obtained from 70 of the 164 patients in our study; 33 of the fresh-frozen samples were from patients with adenocarcinoma, 24 were from patients with SCC, and 13 were from patients with LCC. To study the tumor genomic landscape from different angles, we used targeted NGS to evaluate single-nucleotide variants and short deletions/insertions (indels) in NSCLC samples, including 13 genes, 121 targets, and 372 amplicons associated with these genes. We verified the results with DNA sequencing analysis with an Illumina TruSeq Custom Amplicon Panel (Supplemental Table 1). The genes examined were grouped according to their functions as immune checkpoints (*CD274* [*PD-L1*], *CD276, CTLA4, PDCD1LG2* [*PD-L2*], *LAG3,* and *VTCN1* [*B7H4*]) or EMT genes (*CD44*, *CDH1* [E-cadherin], *TGFβ1, VIM, MMP2, ZEB1,* and *ZEB2*). Libraries were sequenced with an Illumina MiSeq platform using MiSeq v2 (300 cycles). Primary data analysis was performed using the TruSeq Amplicon pipeline (alignment to reference genome hg19) via the Basespace Sequence Hub. Illumina VariantStudio v2.2 software was used for filtering and variant annotation. Variants with frequency lower than 1% in the study population were considered single-nucleotide variants, and variants with frequency higher than 1% were defined as somatic variants. Filtered retained variants had to have a total coverage depth of 100 reads or more and a variant

Table 2. Median Densities of Various Immune Marker–Expressing Cells According to Histologic Tumor Type (N = 164)				
	Median Cell Density, cells/mm^2			
Marker	Adenocarcinoma (n = 94)	SCC (n = 51)	LCC (n = 19)	*P* Value[a]
Panel 1				
MCs (AE1/AE3$^+$)	4780.25	4311.02	4366.56	.55
MCs (AE1/AE3$^+$) PD-L1$^+$	0.39	0.48	1.59	.60
Total CD3$^+$	432.90	246.54	423.12	.02
CD3$^+$CD8$^+$	112.71	68.56	107.55	.04
CD3$^+$PD-1$^+$	59.21	23.23	79.50	.01
CD68$^+$	82.09	69.76	100.60	.36
CD68$^+$PD-L1$^+$	2.03	1.11	3.50	.85
Panel 2				
MCs (AE1/AE3$^+$)	5294.40	5407.66	6570.75	.57
Total CD3$^+$	881.43	687.63	694.81	.55
CD3$^+$CD45RO$^+$	534.29	342.31	592.97	.06
CD3$^+$CD57$^+$	20.33	30.01	21.69	.60
CD3$^+$CD45RO$^+$FOXP3$^+$	145.80	83.14	307.63	.36
CD20$^+$	45.69	41.14	24.84	.41

Abbreviations: LCC, large cell carcinoma; MCs, malignant cells; PD-L1, programmed death ligand-1; SCC, squamous cell carcinoma.

[a] The Kruskal-Wallis test was used to detect differences in continuous variables between histologic tumor types. A *P* value ≤.05 was considered statistically significant.

allele frequency of at least 5% to be considered of adequate quality to be interpreted. The potential pathogenicity of the allele variants was evaluated with 3 publicly available algorithms (Polyphen2, Sift, ClinVar). The distribution of mutations in EMT and immune checkpoint genes in our study cohort was compared with data from The Cancer Genome Atlas. All mutations were tested for correlations with patients' clinicopathologic characteristics, TAIC levels, and outcomes.

Statistical Analysis

The χ^2 test or Fisher exact test was used to examine differences in categorical variables, and the Wilcoxon rank sum test and Kruskal-Wallis test were used to detect differences in continuous variables among groups of patients. In addition, a general linear regression model was used to test the relationships between continuous variables and several other variables, and the residuals were examined to ensure that they were approximately normally distributed. The overall survival (OS) distributions for the patients were estimated using the Kaplan-Meier method. Overall survival was defined as the interval from surgery to death and was estimated using the Kaplan-Meier method, and the log-rank test was used to compare survival among the groups of interest. Regression analysis of the OS data was performed using the Cox proportional hazards model. Variables shown by univariate analysis to be significantly associated with survival were entered into a Cox proportional hazards regression model for multivariate analysis. The statistical software programs SPSS (version 22, IBM, Armonk, New York) and S-Plus (version 8.04, TIBCO, Palo Alto, California) were used to perform the analyses. *P* values less than or equal to .05 were deemed statistically significant.

RESULTS

Clinicopathologic Characteristics

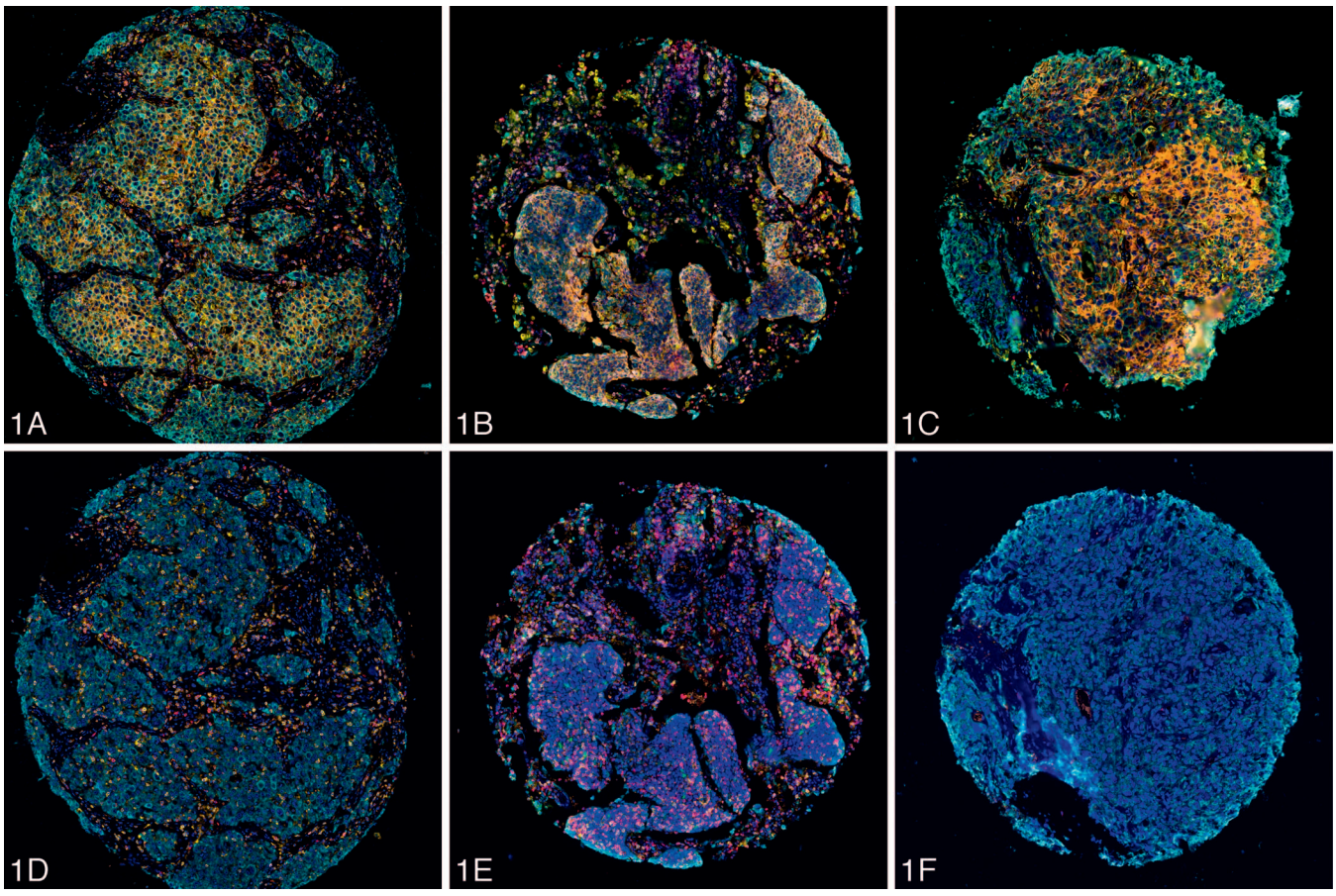
The clinical characteristics of the 164 patients in our cohort are summarized in Table 1 by histologic tumor type. Similar distributions of age and sex were found among the histologic types. Of the 106 patients with a history of tobacco smoking, 51 (48%) had adenocarcinoma, 48 (45%) had SCC, and 7 (7%) had LCC. The median follow-up duration was 43 months (range, 6–180 months). During the follow-up period, disease progression (local recurrence and distant metastasis to liver, bones, lungs, or brain) occurred in 30 of the 164 patients: 16 (53%) with adenocarcinoma, 12 (40%) with SCC, and 2 (7%) with LCC. All patients with distant metastases also had lymph node metastases. Of the 164 patients, 34 (21%) received adjuvant therapy, and 56 (34%) had died of their disease by last follow-up.

PD-L1 Expression in MCs and Macrophages

The median cell densities of MCs and macrophages by histologic tumor type are shown in Table 2. Image analysis and mIF revealed no significant differences in the median densities of PD-L1$^+$ MCs among the histologic types: the median density was 0.39 cells/mm^2 for adenocarcinoma, 0.48 cells/mm^2 for SCC, and 1.59 cells/mm^2 for LCC (P = .60) (Table 2). Using the median percentage of MCs expressing PD-L1 across all the samples as a cutoff, we identified 45 of 94 (47%) PD-L1$^+$ adenocarcinoma samples, 25 of 51 (49%) PD-L1$^+$ SCC samples, and 6 of 19 (31%) PD-L1$^+$ LCC samples (Figure, A through F). For all 3 histologic types, the density of CD68$^+$PD-L1$^+$ TAMs was higher than that of PD-L1$^+$ MCs.

Clinicopathologic Correlations With PD-L1 and TAICs

Certain immune cell phenotypes were significantly associated with patients' clinicopathologic characteristics, as shown in Supplemental Table 2. A general linear regression model demonstrated that low PD-L1 expression on MCs was significantly associated with tobacco use (P = .04) and adenocarcinoma (P = .04). Low density of natural killer T cells (CD3$^+$CD57$^+$) was also associated with tobacco use (P = .04). Low density of cytotoxic T cells (CD3$^+$CD8$^+$) was associated with LCC (P = .04). In addition, low density of memory T cells (CD3$^+$CD45RO$^+$) was associated with



Multiplex immunofluorescence images of representative non–small cell lung cancer core tumor sections analyzed for panel 1 (pancytokeratin, cyan; PD-L1, orange; CD3, red; CD8, pink; PD-1, green; CD68, yellow; DAPI, blue) and panel 2 (pancytokeratin, cyan; CD3, red; CD8, pink; CD45RO, yellow; GranzymeB, orange; FOXP3, green; DAPI, blue) markers. The images reflect the variations in cell phenotypes observed in adenocarcinoma (A and D, panels 1 and 2, respectively), squamous cell carcinoma (B and E, panels 1 and 2, respectively), and large cell carcinoma (C and F, panels 1 and 2, respectively), especially in regard to PD-L1 expression by malignant cells (original magnification ×200).

female sex ($P = .01$) and age 65 years or older ($P = .04$). Importantly, high density of antigen-experienced T cells ($CD3^+PD\text{-}1^+$) was associated with brain metastasis ($P = .04$), and, curiously, high density of B cells ($CD20^+$) was significantly more common among patients who did not receive adjuvant therapy than among those who did ($P = .04$). All these associations remained statistically significant after allowing for the contribution of other clinicopathologic characteristics, as determined by multivariate analysis.

Characterization of Tumor Immune Microenvironments

We observed significantly higher densities of $CD3^+$ T cells ($P = .02$), $CD3^+CD8^+$ cytotoxic T cells ($P = .04$), and $CD3^+PD\text{-}1^+$ antigen–experienced T cells ($P = .01$) in adenocarcinoma and LCC than in SCC (Table 2). No other statistically significant differences were found among the histologic types.

By examining the PD-L1 expression of MCs in conjunction with the density of $CD3^+$ TILs, as proposed by Teng and colleagues,[20] among the different histologic types, we identified all 4 types of tumor microenvironments in our NSCLC samples (Table 3). Type I, adaptive immune resistance ($PD\text{-}L1^+/TIL^+$), was observed in 11 of 94 (12%) adenocarcinoma samples, 7 of 51 (14%) SCC samples, and 4 of 19 (21%) LCC samples. Type II, immunologic ignorance ($PD\text{-}L1^-/TIL^-$), was found in 39 of 94 (41%) adenocarcino-

Table 3. Distribution of Types of Immune Microenvironments in 164 Non–Small Cell Lung Cancer Specimens[a]

PD-L1[b]	TILs[c]	
	Negative	Positive
Adenocarcinoma, No. (%) (n = 94)		
Negative	39 (41)	41 (43)
Positive	3 (3)	11 (12)
Squamous cell carcinoma, No. (%) (n = 51)		
Negative	28 (55)	12 (24)
Positive	4 (8)	7 (14)
Large cell carcinoma, No. (%) (n = 19)		
Negative	7 (37)	5 (26)
Positive	3 (16)	4 (21)

Abbreviations: PD-L1, programmed death ligand-1; TIL, tumor-infiltrating lymphocyte.

[a] The tumor immune microenvironments are based on the percentage of malignant cells expressing PD-L1 and the density of $CD3^+$ TILs.

[b] PD-L1 expression was considered negative if the percentage of malignant cells expressing PD-L1 on the cell membrane was less than or equal to the median percentage for that histologic type and positive if greater than the median.

[c] TILs were considered negative if the density of $CD3^+$ cells was less than or equal to the median density for that histologic type and positive if greater than the median.

ma samples, 28 of 51 (55%) SCC samples, and 7 0f 19 (37%) LCC samples. Type III, intrinsic induction (PD-L1$^+$/TIL$^-$), was found in 3 of 94 (3%) adenocarcinoma samples, 4 of 51 (88%) SCC samples, and 3 of 19 (16%) LCC samples. Type IV, tolerance (PD-L1$^-$/TIL$^+$), was found in 41 of 94 (42%) adenocarcinoma samples, 12 of 51 (24%) SCC samples, and 5 of 19 (26%) LCC samples.

Correlations Between Genomic Profiles, Clinicopathologic Characteristics, and Immune v

Using NGS, we examined the occurrence of variants in several cancer-related genes in fresh-frozen lung cancer tissue specimens obtained from 70 of the 164 NSCLC patients included in our study, and we correlated the mutation status of these genes with patients' clinicopathologic characteristics and immune cell populations. The distribution of mutations in EMT and immune checkpoint genes in our study cohort was compared with data from The Cancer Genome Atlas (Supplemental Table 3). We found similar frequencies of mutations in immune checkpoint and EMT genes in our NSCLC cohort and the Cancer Genome Atlas cohort. The most frequently mutated genes in our cohort were *ZEB2* (n = 10; 14%), *MMP2* (n = 4; 6%), *ZEB1* (n = 2; 3%), *CDH1* (n = 2; 3%), and *CD44* (n = 2; 3%), all of them involved in EMT. On multivariate analysis, *CD44* mutations were significantly associated with tobacco use ($P \leq .05$). The small number of mutations in *TGFβ1* and *VIM* did not allow for multivariate analysis to test for associations with clinicopathologic characteristics.

The genotype and allele frequencies of 5 variants, rs7854303 (*PDCD1LG2* [*PD-L2*]), rs2297136 (*CD274* [*PD-L1*]), rs870849 (*LAG3*), rs10754339 (*VTCN1* [*B7H4*]), and rs231775 (*CTLA4*), in our cohort are summarized in Table 4. Only genotype CC and the C allele of *PD-L2* variant rs7854303 were identified in all patients. On multivariate analysis, the only significant difference in clinicopathologic characteristics by genotype was that the AA genotype of the *CTLA4* rs231775 variant was more common in stage II and III than in stage I tumors ($P = .02$) (Supplemental Table 2). No association was found between rs2297136 (*PD-L1*), rs870849 (*LAG3*), or rs10754339 (*VTCN1*) and any clinicopathologic characteristics.

Our investigation of the association between genomic profiles and immune cell profiles revealed significant correlations between high density of CD3$^+$PD-1$^+$ antigen–experienced T cells and the G allele (GG$^+$GA) of *CTLA4* rs231775 ($R = 0.33$; $P = .01$), high density of CD3$^+$ TILs and *ZEB1* mutation ($R = 0.27$; $P = .04$), and high density of CD3$^+$CD8$^+$ cytotoxic T cells and *ZEB1* mutation ($R = 0.34$; $P = .01$) (Supplemental Table 4).

Survival Analysis

Preliminary examination of Kaplan-Meier survival curves (data not shown) demonstrated that in this study, patients with pathologic disease stages II and III had approximately the same hazard for survival, with a median survival time of 40 months for both groups. Thus, we coded overall pathologic stage as a single dummy variable with a value of 0 for stage I and a value of 1 for stages II and III. The results of the Cox proportional hazards regression models appear in Tables 4 and 5.

The following clinicopathologic variables were significantly associated with poor OS: female sex, tobacco use, tumor size greater than 4.5 cm, N1 status, tumor stage higher than II, SCC histologic type, and absence of adjuvant treatment (Table 5). We also found that lower-than-median densities of CD3$^+$CD8$^+$, CD3$^+$CD45RO$^+$, and CD3$^+$CD45RO$^+$FOXP3$^+$ T cells in the tumor microenvironment predicted shorter OS, as did *ZEB2*-mutant tumors. In addition, *CTLA4* variant rs231775 with AA genotype was associated with shorter OS, whereas allele G was significantly associated with longer OS (adjusted hazard ratio = 2.091, 95% CI = 0.233–2.219, $P = .05$) (Table 4). No other variables included in the multivariate analysis were significantly associated with survival.

Multivariate analysis controlling for CD3$^+$CD8$^+$ T cell density, CD3$^+$CD45RO$^+$ T cell density, *CTLA4* variants, and *ZEB2* mutations revealed that tumor size, N stage, histologic type, and adjuvant treatment significantly predicted survival time. However, when CD3$^+$CD45RO$^+$ T cells, *CTLA4* rs231775 G allele carrier genotype, and *ZEB2* wild type were included as covariates, their relationship to survival was much stronger. Whereas the overall likelihood ratio of the Cox model using tumor size, N staging, histologic type, and treatment was just 21.83, the likelihood ratio with stage, CD3$^+$CD45RO$^+$ T cells, *CTLA4* variant rs231775 G allele carrier genotype, and *ZEB2* wild-type tumors was 31.18.

DISCUSSION

In the present study, we explored the quantitative relationships among TAIC density and EMT and immune checkpoint genomic profiles in 164 patients with nonmetastatic NSCLC. Our results indicate that incorporating genomic profile data into quantitative immune profiling produced several important findings. First, our results validate mIF for immune profiling in NSCLC. Second, we discovered associations between clinicopathologic characteristics of NSCLC and PD-L1 expression by MCs and immune cells. Third, our analyses revealed that higher densities of antigen-experienced T cells were associated with metastasis. Fourth, we found that microenvironment types II (immunologic ignorance) and IV (tolerance) were the most frequent microenvironments in the tested NSCLC tissues. Fifth, we found that patients with early-stage disease, carriers of the G allele of *CTLA4* variant rs231775, patients with wild-type *ZEB2*, and patients with a high density of CD3$^+$CD45RO$^+$ T cells had longer OS than other patients. Finally, we found that *ZEB2* mutations were indicative of EMT and served as a proxy for noninflammatory immune phenotypes (immunologic ignorance and immune tolerance). This information may be important for predicting response to immune checkpoint blockade, particularly in clinical settings where NGS is available but immune profiling is not.

To better understand the role of immune and genomic profiling in early-stage NSCLC, our study was carried out in a cohort of patients with nonmetastatic disease. Our investigation included adenocarcinoma, SCC, and LCC, and, to our knowledge, it is the first study to compare genetic and immune cell profiles among the 3 major histologic types of NSCLC and to determine their impact on patient outcomes.

Using mIF and image analysis, we found that PD-L1 expression by MCs and other cells was associated with several clinicopathologic characteristics and with disease outcomes. We found no differences in PD-L1 expression on MCs among the histologic tumor types, but our results showing that 31.51% of LCC samples had positive PD-L1 expression on MCs agreed with those of several studies of large cell neuroendocrine carcinoma of the lung.[22,23] This

Table 4. Distribution of Genetic Variants in 70 Patients With Non–Small Cell Lung Cancer and Association of These Variants With Overall Survival

Gene	Chromosome Position	Genetic Variant	Genotype	No. (%) of Cases	Univariate Analysis[a]			Multivariate Analysis[b]	
					HR (95% CI)	HR	*P* Value	HR (95% CI)	*P* Value
PDCD1LG2 (*PD-L2*)	9p24.1	rs7854303	TT	0 (0)	1.0 (reference)				
			TC	0 (0)					
			CC	70 (100)					
			C allele carrier	70 (100)	—				
CD274 (*PD-L1*)	9p24.1	rs2297136	GG	15 (21)	1.0 (reference)				
			GA	25 (36)					
			AA	30 (43)					
			A allele carrier	55 (79)	0.57 (0.20–1.63)	−0.056	.29		
LAG3	12p13.31	rs870849	TT	9 (13)	1.0 (reference)				
			TC	32 (46)					
			CC	29 (41)					
			C allele carrier	61 (87)	2.65 (0.73–9.57)	0.97	.13	0.18 (0.00–0.98)	.42
VTCN1 (*B7H4*)	1p13.1-p12	rs10754339	GG	4 (6)	1.0 (reference)				
			GA	13 (19)					
			AA	53 (76)					
			A allele carrier	66 (94)	1.55 (0.20–11.54)	0.44	.67		
CTLA4	2q33.2	rs231775	AA	34 (49)	1.0 (reference)				
			AG	26 (37)					
			GG	10 (14)					
			G allele carrier	36 (51)	0.22 (0.25–0.99)	−1.48	.04	2.091 (0.233–2.219)	.05

Abbreviation: HR, hazard ratio (β coefficient).

[a] Univariate analysis was carried out without any adjustments in order to generate hazard ratios for death with confidence intervals for each of the genotypes.
[b] Multivariate analysis was carried out to analyze the effects of several risk genotypes on death.

Table 5. Univariate and Multivariate Cox Proportional Hazards Models of Associations of Overall Survival With Clinicopathologic Characteristics and Immune Cell Profiles

	Univariate Analysis[a]			Multivariate Analysis[b]	
	HR (95% CI)	HR	*P*	HR (95% CI)	*P*
Clinicopathologic characteristics					
Age, y, >65 versus ≤65	0.99 (0.97–1.01)	−0.005	.65		
Sex					
Female	2.18 (1.29–3.45)	0.75	.003		
History of tobacco use					
Yes	1.01 (1.009–1.016)	0.009	.007		
Tumor size, cm					
>4.5	1.23 (1.11–1.37)	0.21	<.001		
Lymph node (N) status					
N0	0.387 (0.219–0.684)	−0.95	<.001		
N1 (reference)					
Recurrence-free	0.77 (0.47–1.25)	−0.26	.29		
Stage					
I	0.369 (0.225–0.605)	−0.997	<.001	0.437 (0.243–0.785)	.006
>II (reference)					
Histologic type					
Adenocarcinoma	2.782 (0.857–9.037)	1.023	.08		
SCC	3.231 (0.971–10.747)	1.173	.05	2.989 (1.123–7.954)	.02
LCC (reference)					
Adjuvant therapy					
Yes	1.724 (1.038–2.864)	−0.545	.03	0.428 (0.200–0.914)	.02
Immune cells, ≤median versus >median					
MCs PD-L1$^+$/mm^2					
>1.03	1.015 (0.589–1.750)	0.015	.95		
CD3$^+$ cells/mm^2					
>327.33	1.077 (0.625–1.857)	0.075	.78		
CD3$^+$CD8$^+$ cells/mm^2					
>110.35	1.744 (1.003–3.031)	0.556	.04		
CD3$^+$PD-1$^+$ cells/mm^2					
>40.36	1.264 (0.731–2.186)	0.234	.40		
CD68$^+$ cells/mm^2					
>77.81	1.125 (0.653–1.939)	0.118	.67		
CD68$^+$PD-L1 cells/mm^2					
>2.54	1.234 (0.715–2.129)	0.210	.45		
CD3$^+$CD45RO$^+$ cells/mm^2					
>562.45	1.716 (0.984–2.993)	0.540	.05	2.265 (1.276–4.022)	.005
CD3$^+$CD45RO$^+$FOXP3$^+$ cells/mm^2					
>126.84	1.923 (1.090–3.396)	0.654	.02		
CD3$^+$CD57$^+$ cells/mm^2					
>22.07	1.085 (0.624–1.884)	0.081	.77		
CD20$^+$ cells/mm^2					
EMT gene mutations (% of samples)					
CD44 (2.8)	0.899 (0.666–1.208)	0.573	.44		
CDH1 (2.8)	1.414 (0.082–2.357)	0.345	.80		
MMP2 (5.7)	0.091 (0.022–0.459)	−2.409	.64		
ZEB1 (2.8)	1.166 (0.155–8.659)	0.147	.88		
ZEB2 (2.8)	0.151 (0.011–1.207)	−1.891	.04	−1.429 (0.240–1.972)	.04

Abbreviations: EMT, epithelial-mesenchymal transition; HR, hazard ratio (β coefficient); LCC, large cell carcinoma; MC, malignant cell; PD-L1, programmed death ligand-1; SCC, squamous cell carcinoma.

[a] Univariate analysis was carried out without any adjustments to generate hazard ratios for overall survival with confidence intervals for each parameter.

[b] Multivariate analysis was carried out to analyze the effects of several risk parameters on overall survival.

similarity suggests that patients with the large cell subtype of NSCLC may benefit from checkpoint inhibitor therapies. In agreement with findings reported by Calles and colleagues,[24] we found that high expression of PD-L1 on MCs was associated with tobacco use and the adenocarcinoma histologic type. Interestingly, we found that low densities of natural killer T cells were also associated with tobacco use; this finding was in concordance with Hogan and colleagues'[25] study, which showed a significant reduction of natural killer cells in tobacco users. That reduction was accompanied by significant defects in cytokine production, suggesting that depletion of natural killer T cells can interfere with antitumor immune responses.

Similarly, our study showed that low densities of cytotoxic T cells were associated with LCC and that low densities of memory T cells were associated with female sex and older age, suggesting that alterations in the regulation of the immune system can increase the likelihood of tumor invasion and progression. As suggested by Prado-Garcia and colleagues,[26] alterations in the immune system induced by tumor cells can lead to T-cell dysfunction and reduce T cells' ability to attack tumors. It is also known that increasing age can produce downregulation of immune cells.[27]

Interestingly, shorter survival time was associated with high densities of $CD3^+PD\text{-}1^+$ antigen–experienced T cells, suggesting that PD-1 plays an important role in facilitating NSCLC metastasis. A similar finding was made by Wang et al,[28] who found that, in cervical cancer, the interaction between PD-1 and PD-L1 can trigger immune responses and facilitate tumor growth and metastasis. Most, if not all, malignancies trigger different immune responses through intricate interactions between tumor cells and the host's immune cells,[29] suggesting that the progression of cancer is influenced by complex tumor–3immune cell interactions driven by characteristics of the tumor, the behavior of inflammatory cells, and genetic mutations.

Our study also identified differences in the immune profiles of the histologic types of NSCLC that may affect the choice of therapy. Densities of different subpopulations of T cells, including cytotoxic T cells, were higher in adenocarcinoma and LCC than in SCC; these results agreed with those of our previous study.[30] In addition, we also observed higher densities of $CD3^+PD\text{-}1^+$ antigen–experienced T cells in adenocarcinoma and LCC than in SCC, suggesting that adenocarcinoma and LCC provoke more immune exhaustion and, thus, that immunotherapeutic intervention may be warranted for these 2 histologic types. Previous studies have shown that activation of the PD-1 pathway mediates inhibition of T cells and that PD-1 expression is an independent predictive factor for prognosis.[31] Using the tumor microenvironment criteria proposed by Teng and colleagues,[20] we determined that the most common microenvironments in our cohort were immunologic ignorance ($PD\text{-}L1^-/TIL^-$) and immunologic tolerance ($PD\text{-}L1^-/TIL^+$), suggesting that complementary strategies, as proposed by Teng et al,[20] will be necessary to induce a response to immunotherapy in patients with these microenvironments.

Our findings also highlighted the role of EMT in genomic profiling to predict outcomes in NSCLC. The EMT is activated in cancer cells and is induced by transcription factors such as Snail, Twist, and ZEB1/ZEB2 to modify the transcriptional machinery, alter translation and protein stability, and promote invasion and metastasis.[32] In fact, we demonstrated that *ZEB1* mutations were inversely correlated with levels of $CD3^+$ and $CD3^+CD8^+$ T cells, suggesting a molecular link between EMT and immunosuppression, 2 key drivers of cancer progression. We also found that mutations in the EMT-associated gene *CD44* were associated with adenocarcinoma and SCC and with tobacco use, in agreement with other studies.[32–34] Interestingly, tobacco use induces metaplastic bronchial squamous epithelium that exhibits increased hyaluronan and CD44 expression in the proliferating basal layers. In premalignant bronchial dysplasia, the entire epithelial thickness shows aberrant hyaluronan-CD44 expression, indicating that squamous malignant transformation is closely associated with CD44 expression and with tobacco use.[35] We also found that *CD44*-mutant tumors were associated with high densities of $CD20^+$ B cells, which suggests that CD44 also may participate in several other immune-related processes, as previously demonstrated by Zittermann and colleagues.[36]

Over the last decade, significant progress has been made in systemic therapies to improve the length of good-quality survival in patients with metastatic NSCLC. However, many questions remain about the behavior of nonmetastatic NSCLC. One such issue relates to the expected 5-year survival rate, which ranges from 36% to 92% for patients with early-stage (ie, stage I, II, or IIIA) disease.[3] Another question relates to current studies of immunotherapy in the context of nonmetastatic NSCLC.[3] Not all NSCLC patients have a response to immunotherapy because the heterogeneity of immune cell profiles in NSCLC tumors probably depends on other factors, for instance EMT and immune checkpoint genes' control of cancer progression.[7,37] In this context, in our cohort, multivariate analysis reliably predicted longer OS for patients with a high density of $CD3^+CD45RO^+$ memory T cells and those with the G allele of *CTLA4* variant rs231775 and shorter OS when nonmetastatic NSCLC tumors have *ZEB2* mutations. In fact, the presence of a high density of $CD3^+CD45RO^+$ immune cells in the tumor region is correlated with favorable clinical outcomes in epithelial lung cancers.[38] In addition, several studies have assessed the contribution of *CTLA4* polymorphisms in the context of immune checkpoint inhibition.[39–44] Our multivariate Cox analysis revealed that the *CTLA4* rs231775 AA genotype is an adverse prognostic indicator for NSCLC patients. Although associations between CTLA4 and outcomes have been reported, these lacked independent validation. As shown by Deng and colleagues,[44] *CTLA4* expression was not significantly prognostic in their cohort of 1432 patients on univariate analysis but was significant in their cohort of patients with available data for multivariate analysis.

Our analysis has a number of limitations that could not be addressed in this first quantitative study. The first of these limitations is the small cohort, which included 164 retrospectively collected cases. However, the maximum follow-up time was 168 months, the robustness of clinical variables was ascertained, and survival information was available. A second limitation is that we included only 70 tissue samples in our genomic analysis. However, a preliminary examination of Kaplan-Meier survival curves (data not shown) demonstrated that patients with pathologic stages II and III disease had approximately the same hazard for survival, with a median survival time of 40 months for both groups. A third limitation of the study concerns the noninclusion of normal tissue or blood

samples for the genomic analysis, which may have affected the allele frequency determination. A fourth limitation was the tissue microarray format we used, which could have induced underrepresentation or overrepresentation of PD-L1 due to tumor heterogeneity and the small area of analysis. Fifth, we lacked available data regarding responses to anti–PD-1/anti–PD-L1 monoclonal antibodies in our patient population. Sixth, even with accumulating evidence supporting the contribution of germline genetics to host immunity, knowledge about host genetic factors as predictive biomarkers of clinical outcomes is limited. Moreover, at present, no systematic study of genetic variants as surrogates for immunotherapy outcomes at the genome-wide level is available. Larger study cohorts will be required for the discovery of low-penetrance germline loci associated with immunotherapy outcomes. This clearly indicates the need for a large international collaboration pooling patient resources. The implementation of genome-wide association studies coupled with large-scale immune phenotyping will help to identify common genetic variations associated with a large population–based scale.[42] Finally, mIF can be used to simultaneously identify specific proteins on a single slide. This technique allows study of the immune contexture using paraffin-embedded tumor tissues. In addition, multiplexed image analysis methods are highly advantageous for investigating immune evasion mechanisms and for discovering potential biomarkers to assess mechanisms of action and to predict response to a given treatment.[45,46] The combination of NGS data with mIF may facilitate disease management and the appropriate application of immunotherapeutic agents.

In conclusion, incorporating a genomic profile into quantitative mIF revealed a variety of factors associated with the behavior of nonmetastatic NSCLC, including specific clinicopathologic characteristics and immunomodulatory control of local disease progression. Thus, these tools might be useful for predicting whether patients with early-stage, nonmetastatic NSCLC will benefit from combinations of targeted therapy, chemotherapy, and immunotherapy.

Editorial support was provided by Amy Ninetto, PhD, ELS, of Scientific Publications, Research Medical Library, The University of Texas MD Anderson Cancer Center.

References

1. Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM Classification for Lung Cancer. *J Thorac Oncol*. 2016; 11(1):39–51.

2. Ghysen K, Vansteenkiste J. Immunotherapy in patients with early stage resectable nonsmall cell lung cancer. *Curr Opin Oncol*. 2019;31(1):13–17.

3. Deslypere G, Gullentops D, Wauters E, Vansteenkiste J. Immunotherapy in non-metastatic non-small cell lung cancer: can the benefits of stage IV therapy be translated into earlier stages? *Ther Adv Med Oncol*. 2018;10: 1758835918772810.

4. Yu H, Boyle TA, Zhou C, et al. PD-L1 expression in lung cancer. *J Thorac Oncol*. 2016;11:964–975.

5. Sun JM, Zhou W, Choi YL, et al. Prognostic significance of PD-L1 in patients with non-small cell lung cancer: a large cohort study of surgically resected cases. *J Thorac Oncol*. 2016;11(7):1003–1011.

6. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1–patients with non-small cell lung cancer: a large cohort study of surgically resected cases. *N Engl J Med*. 2016;375:1823–1833.

7. Kim S, Koh J, Kim MY, et al. PD-L1 expression is associated with epithelial-to-mesenchymal transition in adenocarcinoma of the lung. *Hum Pathol*. 2016;58: 7–14.

8. Marwitz S, Scheufele S, Perner S, Reck M, Ammerpohl O, Goldmann T. Epigenetic modifications of the immune-checkpoint genes CTLA4 and PDCD1 in non-small cell lung cancer results in increased expression. *Clin Epigenetics*. 2017;9:51.

9. Papanicolau-Sengos A, Yang Y, Pabla S, et al. Identification of targets for prostate cancer immunotherapy. *Prostate*. 2019;79(5):498–505.

10. O'Leary K, Shia A, Schmid P. Epigenetic regulation of EMT in non-small cell lung cancer. *Curr Cancer Drug Target*s. 2018;18(1):89–96.

11. Chae YK, Chang S, Ko T, et al. Epithelial-mesenchymal transition (EMT) signature is inversely associated with T-cell infiltration in non–small cell lung cancer (NSCLC). *Sci Rep*. 2018;8(1):2918.

12. Chen L, Gibbons DL, Goswami S, et al. Metastasis is regulated via microRNA-200/ZEB1 axis control of tumour cell PD-L1 expression and intratumoral immunosuppression. *Nat Commun*. 2014;5:5241.

13. Kurimoto R, Ebata T, Iwasawa S, et al. Pirfenidone may revert the epithelial-to-mesenchymal transition in human lung adenocarcinoma. *Oncol Lett*. 2017; 14(1):944–950.

14. Alsuliman A, Colak D, Al-Harazi O, et al. Bidirectional crosstalk between PD-L1 expression and epithelial to mesenchymal transition: significance in claudin-low breast cancer cells. *Mol Cancer*. 2015;14:149.

15. Ma Y, Adjemian S, Mattarollo SR, et al. Anticancer chemotherapy-induced intratumoral recruitment and differentiation of antigen-presenting cells. *Immunity*. 2013;38(4):729–741.

16. Mazzaschi G, Madeddu D, Falco A, et al. Low PD-1 expression in cytotoxic CD8(+) tumor-infiltrating lymphocytes confers an immune-privileged tissue microenvironment in NSCLC with a prognostic and predictive value. *Clin Cancer Res*. 2018;24(2):407–419.

17. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of Lung Tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol*. 2015;10(9): 1243–1260.

18. Parra ER, Uraoka N, Jiang M, et al. Validation of multiplex immunofluorescence panels using multispectral microscopy for immune-profiling of formalin-fixed and paraffin-embedded human tumor tissues. *Sci Rep*. 2017; 7(1):13380.

19. Gorris MAJ, Halilovic A, Rabold K, et al. Eight-color multiplex immunohistochemistry for simultaneous detection of multiple immune checkpoint molecules within the tumor microenvironment. *J Immunol*. 2018;200(1): 347–354.

20. Teng MW, Ngiow SF, Ribas A, Smyth MJ. Classifying cancers based on T-cell infiltration and PD-L1. *Cancer Res*. 2015;75(11):2139–2145.

21. Machado-Rugolo J, Fabro AT, Ascheri D, et al. Usefulness of complementary next-generation sequencing and quantitative immunohistochemistry panels for predicting brain metastases and selecting a treatment outcomes of non-small cell lung cancer. *Hum Pathol*. 2019;83:177–191.

22. Eichhorn F, Harms A, Warth A, Muley T, Winter H, Eichhorn ME. PD-L1 expression in large cell neuroendocrine carcinoma of the lung. *Lung Cancer*. 2018;118:76–82.

23. Tsuruoka K, Horinouchi H, Goto Y, et al. PD-L1 expression in neuroendocrine tumors of the lung. *Lung Cancer*. 2017;108:115–120.

24. Calles A, Liao X, Sholl LM, et al. Expression of PD-1 and its ligands, PD-L1 and PD-L2, in smokers and never smokers with KRAS-mutant lung cancer. *J Thorac Oncol*. 2015;10(12):1726–1735.

25. Hogan AE, Corrigan MA, O'Reilly V, et al. Cigarette smoke alters the invariant natural killer T cell function and may inhibit anti-tumor responses. *Clin Immunol*. 2011;140(3):229–235.

26. Prado-Garcia H, Romero-Garcia S, Aguilar-Cazares D, Meneses-Flores M, Lopez-Gonzalez JS. Tumor-induced CD8+ T-cell dysfunction in lung cancer patients. *Clin Dev Immunol*. 2012;2012:741741.

27. Fulop T, Larbi A, Kotb R, de Angelis F, Pawelec G. Aging, immunity, and cancer. *Discov Med*. 2011;11(61):537–550.

28. Wang S, Li J, Xie J, et al. Programmed death ligand 1 promotes lymph node metastasis and glucose metabolism in cervical cancer by activating integrin beta4/SNAI1/SIRT3 signaling pathway. *Oncogene*. 2018;37(30):4164–4180.

29. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. *Nature*. 2008;454(7203):436–444.

30. Parra ER, Behrens C, Rodriguez-Canales J, et al. Image analysis-based assessment of PD-L1 and tumor-associated immune cells density supports distinct intratumoral microenvironment groups in non-small cell lung carcinoma patients. *Clin Cancer Res*. 2016;22(24):6278–6289.

31. Paulsen EE, Kilvaer TK, Khanehkenari MR, et al. Assessing PDL-1 and PD-1 in non-small cell lung cancer: a novel immunoscore approach. *Clin Lung Cancer*. 2017;18(2):220–233 e8.

32. Gonzalez DM, Medici D. Signaling mechanisms of the epithelial-mesenchymal transition. *Sci Signal*. 2014;7(344): re8.

33. Leung EL, Fiscus RR, Tung JW, et al. Non-small cell lung cancer cells expressing CD44 are enriched for stem cell-like properties. *PLoS One*. 2010; 5(11):e14062.

34. Chou YE, Hsieh MJ, Chiou HL, Lee HL, Yang SF, Chen TY. CD44 gene polymorphisms on hepatocellular carcinoma susceptibility and clinicopathologic features. *Biomed Res Int*. 2014;2014:231474.

35. Penno MB, August JT, Baylin SB, et al. Expression of CD44 in human lung tumors. *Cancer Res*. 1994;54(5):1381–1387.

36. Zittermann SI, Achino BI, Agriello EE, Halperin N, Ramhorst RE, Fainboim L. Modulation of CD44 in acute lymphoblastic leukemia identifies functional and phenotypic differences of human B cell precursors. *Eur J Haematol*. 2001;66(6): 377–382.

37. Koh J, Go H, Keam B, et al. Clinicopathologic analysis of programmed cell death-1 and programmed cell death-ligand 1 and 2 expressions in pulmonary

adenocarcinoma: comparison with histology and driver oncogenic alteration status. *Mod Pathol*. 2015;28(9):1154–1166.

38. Sheng SY, Gu Y, Lu CG, Zou JY, Hong H, Wang R. The distribution and function of human memory T cell subsets in lung cancer. *Immunol Res*. 2017; 65(3):639–650.

39. Breunis WB, Tarazona-Santos E, Chen R, Kiley M, Rosenberg SA, Chanock SJ. Influence of cytotoxic T lymphocyte-associated antigen 4 (CTLA4) common polymorphisms on outcome in treatment of melanoma patients with CTLA-4 blockade. *J Immunother*. 2008;31(6):586–590.

40. Queirolo P, Morabito A, Laurent S, et al. Association of CTLA-4 polymorphisms with improved overall survival in melanoma patients treated with CTLA-4 blockade: a pilot study. *Cancer Invest*. 2013;31(5):336–345.

41. Hamid O, Schmidt H, Nissan A, et al. A prospective phase II trial exploring the association between tumor microenvironment biomarkers and clinical activity of ipilimumab in advanced melanoma. *J Transl Med*. 2011;9:204–220.

42. Kirchhoff T, Ferguson R. Germline genetics in immuno-oncology: from genome-wide to targeted biomarker strategies. *Methods Mol Biol*. 2020;2055:93–117.

43. Lim YW, Chen-Harris H, Mayba O, et al. Germline genetic polymorphisms influence tumor gene expression and immune cell infiltration. *Proc Natl Acad Sci U S A*. 2018;115:E11701–E11710.

44. Deng L, Gyorffy B, Na F, et al. Association of PDCD1 and CTLA-4 gene expression with clinicopathological factors and survival in non-small-cell lung cancer: results from a large and pooled microarray database. *J Thorac Oncol*. 2015;10(7):1020–1026.

45. Hofman P, Badoual C, Henderson F, et al. Multiplexed immunohistochemistry for molecular and immune profiling in lung cancer—just about ready for prime-time? *Cancers (Basel)*. 2019;11(3):283.

46. Parra ER, Francisco-Cruz A, Wistuba II. State-of-the-art of profiling immune contexture in the era of multiplexed staining and digital analysis to study paraffin tumor tissues. *Cancers (Basel)*. 2019;11(2):247.

Pharmacogenomics and Personalized Medicine

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Relevance of PD-L1 Non-Coding Polymorphisms on the Prognosis of a Genetically Admixed NSCLC Cohort

This article was published in the following Dove Press journal:
Pharmacogenomics and Personalized Medicine

Juliana Machado-Rugolo,[1,2] Tabatha Gutierrez Prieto,[1] Alexandre Todorovic Fabro,[3] Edwin Roger Parra Cuentas,[4] Vanessa Karen Sá,[5] Camila Machado Baldavira,[1] Claudia Aparecida Rainho,[6] Erick C Castelli,[7,8] Cecilia Farhat,[1] Teresa Yae Takagaki,[9] Maria Aparecida Nagai,[10,11] Vera Luiza Capelozzi[1]

[1]Laboratory of Genomics and Histomorphometry, Department of Pathology, University of São Paulo Medical School (USP), São Paulo, Brazil; [2]Health Technology Assessment Center, Clinical Hospital (HCFMB), Medical School of São Paulo State University (UNESP), Botucatu, São Paulo, Brazil; [3]Department of Pathology and Legal Medicine, Ribeirão Preto School of Medicine, University of São Paulo (FMRP-USP), Ribeirão Preto, Brazil; [4]Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA; [5]Laboratory of Genomics and Molecular Biology, Centro Internacional De Pesquisa (CIPE), AC Camargo Cancer Center, São Paulo, SP, Brazil; [6]Department of Chemical and Biological Sciences, Institute of Biosciences, São Paulo State University (UNESP), Botucatu, São Paulo, Brazil; [7]Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unit (UNIPEX), Medical School of São Paulo State University (UNESP), Botucatu, São Paulo, Brazil; [8]Department of Pathology, Medical School of São Paulo State University (UNESP), Botucatu, São Paulo, Brazil; [9]Division of Pneumology, Heart Institute (Incor), Clinical Hospital, University of São Paulo Medical School (USP), São Paulo, São Paulo, Brazil; [10]Department of Radiology and Oncology, University of São Paulo Medical School (USP), São Paulo, Brazil; [11]Laboratory of Molecular Genetics, Center for Translational Research in Oncology, Cancer Institute of São Paulo (ICESP), São Paulo, Brazil

Correspondence: Vera Luiza Capelozzi
Department of Pathology, University of São Paulo Medical School (USP), Av Dr Arnaldo 455, Room 1143, São Paulo, São Paulo State, 01246-903, Brazil
Tel +55 11 3061 7427
Fax +55 11 3064-2744
Email vera.capelozzi@fm.usp.br

Purpose: Although non-small cell lung cancer (NSCLC) remains a deadly disease, new predictive biomarkers have emerged to assist in managing the disease, of which one of the most promising is the programmed death-ligand 1 (PD-L1). Each, *PD-L1* variant seem to modulate the function of immune checkpoints differently and affect response to adjuvant treatment and outcome in NSCLC patients. We thus investigated the influence of these PD-L1 genetic variations in genetically admixed NSCLC tissue samples, and correlated these values with clinicopathological characteristics, including prognosis.

Materials and Methods: We evaluated PD-L1 non-coding genetic variants and protein expression in lung adenocarcinomas (ADC), squamous cell carcinomas (SqCC), and large cell carcinomas (LCC) in silico. Microarray paraffin blocks from 70 samples of ADC (N=33), SqCC (N=24), and LCC (N=13) were used to create PD-L1 multiplex immunofluorescence assays with a Cell Signaling E1L3N clone. Fifteen polymorphisms of the *PD-L1* gene were investigated by targeted sequencing and evaluated in silico using dedicated tools.

Results: Although *PD-L1* polymorphisms seemed not to interfere with protein expression, PD-L1 expression varied among different histological subtypes, as did clinical outcomes, with the rs4742098A>G, rs4143815G>C, and rs7041009G>A variants being associated with relapse (P=0.01; P=0.05; P=0.02, respectively). The rs7041009 GG genotype showed a significant correlation with younger and alive patients compared to carriers of the A allele (P=0.02 and P<0.01, respectively). The Cox regression model showed that the rs7041009 GG genotype may influence OS (P<0.01) as a co-dependent factor associated with radiotherapy and recurrence in NSCLC patients. Furthermore, the Kaplan–Meier survival curves showed that rs7041009 and rs4742098 might impact PPS in relapsed patients. In silico approaches identified the variants as benign.

Conclusion: *PD-L1* non-coding variants play an important role in modulating immune checkpoint function and may be explored as immunotherapy biomarkers. We highlight the rs7041009 variant, which impacts OS and PPS in NSCLC patients.

Keywords: *PD-L1*, non-small cell lung cancer, single nucleotide polymorphisms, next-generation sequencing

Introduction

Lung cancer is one of the leading causes of cancer-related death globally.[1] Although there are different types of lung cancer, non-small cell lung cancer (NSCLC) represents 85% of all primary lung tumors. NSCLC is a grim disease that is aggravated by the fact that patients normally either receive their diagnosis at advanced stages or present with recurrent disease after initial locoregional

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/PGPM.S286717

treatment.[2] Over the last few decades, conventional chemotherapy, mainly platinum-based chemotherapy, used to be the only therapeutic option for those not eligible for radical intent treatment: a treatment with limited efficacy and very few long-term survivors (5-year overall survival less than 15%). Furthermore, these patients often lacked therapeutic options beyond first-line treatment.[3]

More recently, though, immune checkpoint molecules involved in tumor immune evasion were identified and immune checkpoint inhibitors (ICIs) were introduced in antitumor immunotherapy. This new therapeutic approach targets an inhibitory receptor, the programmed cell death-1 (PD-1) receptor, to assist the immune system in identifying and neutralizing malignant cells. However, tumor cells may evade the host immunosurveillance by expressing the programmed death-1-ligand 1 (PD-L1) as an adaptive, resistant mechanism to suppress this inhibitory receptor.[4] Thus, because PD-L1 up-regulation by tumor cells can protect them from antitumor immune response, the blockade of PD-L1/PD-1 interactions has been recently selected for antitumor immune therapy.[5] Agents targeting the PD-1/PD-L1 signaling pathway have shown promising responses in different types of cancer, including NSCLC. These results point to PD-L1 protein expression as a potential predictive marker for a successful blockade of PD-L1/PD-1 interactions.[6] However, several challenges remain in producing robust evidence to support the use of this biomarker.

In this context, some studies with NSCLC patients have demonstrated that those with more than 50% PD-L1 positive tumor cells are non-responders to anti-PD-1/PD-L1 treatment. In contrast, others have shown that patients whose tumors do not express PD-L1 are good responders.[7–9] To explain the controversies that affect PD-L1 expression, some studies have considered that the heterogeneity between axis expression and response to PD-1/PD-L1 treatment in NSCLC depends on other factors, such as more precise methods to investigate immune evasion mechanisms and the immune microenvironment, as well as greater knowledge on the immune checkpoint genomic profile and the genetic variants of the *PD-L1* gene.[10–13]

Some genetic variants have been shown to affect normal gene activation and transcriptional initiation, and hence influence the amount of mRNA and encoded protein in the cell.[14] Non-coding variants also presumably affect genetic regulatory elements, since a majority of driver variants in cancer genomes occur in non-coding regions.[15,16]

Thus, several studies have investigated the association between *PD-1* and *PD-L1* genetic variants and the risk of various cancers, but their findings have yet failed to completely elucidate this question.[17] A previous study suggested that *PD-L1* polymorphism may predict chemotherapy response and survival rates in advanced-stage NSCLC patients after first-line paclitaxel-cisplatin.[18] More recently, *PD-L1* copy number variations, point mutations, and 3′-UTR disruptions have been highlighted as genetic mechanisms of PD-L1 deregulation.[19] Furthermore, previous research on Brazilian patients suggested that their ethnic background could account for their distinct cancer's molecular profile, perhaps due to their characteristic genetic admixture, inherited from European, African, and Native American ancestors.[20–22]

We hypothesize that *PD-L1* non-coding genetic variants modulate the function of this immune checkpoint in NSCLC. To explore this issue, we investigated fifteen *PD-L1* non-coding genetic variants using next-generation sequencing (NGS) in a Brazilian cohort, aiming to uncover the effect of their ethnic admixture on NSCLC. We also combined our analyses with an in-silico approach to predict the impact of these genetic variants on the disease. We evaluated the associations between PD-L1 protein expression level and clinicopathological characteristics, including the prognosis of NSCLC patients undergoing surgical resection, glimpsing the impact of genetic variants on post progression survival (PPS) and overall survival (OS). Herein, in this context, we intend to expand the existing literature on *PD-L1* gene alterations at the genetic level and their impact on NSCLC in patients from different ethnicities, thus increasing the knowledge about the molecular basis of immunotherapy biomarkers.

Materials and Methods

Cohort

In this retrospective multi-center study, we obtained archival formalin-fixed paraffin-embedded histologic tumor sections from 70 patients diagnosed with NSCLC (33 adenocarcinomas [ADC], 24 squamous cell carcinoma [SqCC] and 13 large cell carcinoma [LCC]) who underwent surgical resection between January 1, 1995, and December 31, 2015. Patients had been treated at the Hospital das Clínicas of the University of São Paulo Medical School (HC-FMUSP),

at the Heart Institute of the University of São Paulo (INCOR), at the Cancer Institute of São Paulo (ICESP), and at the A.C. Camargo Cancer Center in São Paulo, Brazil.

All samples were histologically reviewed by lung pathologists who selected samples with at least 30% of lung cancer cells before nucleic acid extraction. The samples were classified using the 2017 International Association for the Study of Lung Cancer (IASLC) classification system.[23] The clinicopathological features of patients were obtained from the medical records. The study was approved in accordance with the ethical standards of the responsible committee on human experimentation local (Research Ethics Committee of University of São Paulo Medical School - CAAE: 79769017.1.0000.5440; opinion number: 2.673.320) and with the 1964 Helsinki declaration. A waiver of the requirement for informed consent was obtained from committee, and to identity of the subjects under this retrospective analysis was omitted and anonymized.

Multiplex Immunofluorescence Staining

We performed a Multiplex immunofluorescence (mIF) staining using methods that had been previously described and validated.[24,25] Four-micrometer-thick consecutive TMA sections were stained using an automated staining system (BOND-RX; Leica Biosystems, Buffalo Grove, IL) to characterize PD-L1 (clone E1L3N, dilution 1:100; Cell Signaling Technology, Danvers, MA). The PD-L1 marker was stained with its respective fluorophore from the Opal 7 kit (catalogue #NEL797001KT; Akoya Biosciences/PerkinElmer, Waltham, MA). A complete validation using immunofluorescence (IF) allowed us to obtain a uniform, specific, and appropriate signal across all the channel; ie, a well-balanced staining pattern for the multiplex staining.[24,25] We also defined and optimized the correct fluorophore signals between 10 and 30 counts of intensity to maintain good balance and similar thresholds of intensity across all antibodies. In parallel, to detect possible variations in staining and optimize the separation of the signal, positive and negative (autofluorescence) controls were included during the staining process to ensure that all the antibodies performed well together. Autofluorescence controls with an expected spectral resolution of 488nm were able to accurately remove the autofluorescence from all the label signals during the analysis. The stained slides were then scanned using a multispectral microscope, the Vectra Polaris 3.0 imaging system (Akoya Biosciences/PerkinElmer, Waltham, MA), under fluorescence conditions.

Multiplex Immunofluorescence Quantitation

Multispectral images of tumor sections from each core were analyzed with inForm 2.2.1 (Akoya Biosciences/PerkinElmer, Waltham, MA) software Individual cells, which were defined by nuclei staining and identified by the InForm cell segmentation tool, were subjected to a phenotyping pattern-recognition learning algorithm to characterize co-localization of the various cell populations using panel labeling.[26] The panel labeling was as follows: Malignant cells (MCs), with the AE1/AE3+ marker, including those with and without PD-L1 expression (AE1/AE3+ PD-L1+ and AE1/AE3+ PD-L1-, respectively). The individual cell phenotype report produced by the InForm software was processed using Excel 2010 (Microsoft. Houston, TX), and a final summary of the data, which contained the median of each individual phenotype (given as number of cells/mm^2) and the percentage of macrophages and MCs expressing PD-L1, was created for statistical analysis. If the percentage of MCs or macrophages expressing PD-L1was greater than the median value, the PD-L1 expression was considered positive. If the percentage of macrophages or MCs expressing PD-L1 was lower than or equal to the median, the PD-L1 expression was considered negative.

DNA Extraction

Genomic DNA (gDNA) was extracted from frozen NSCLC tissue using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations. DNA concentration was measured using the Qubit® 3.0 Fluorometer (Invitrogen, Life Technologies, CA, USA). DNA integrity was assessed using the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

Sequencing and Data Analysis

We performed a *PD-L1* (*CD274*) full gene screening by deep targeted sequencing using the TruSeq Custom Amplicon Panel v1.5 kit (TSCAP, Illumina, SanDiego, CA) and the MiSeq platform (Illumina, SanDiego, CA). The DNA libraries were performed according to the manufacturer's instructions and consisted of 150 bp paired-end reads (300 cycles).

We performed an NGS data analysis on the Molecular Genetics and Bioinformatics Laboratory of the Experimental Research Unit (UNIPEX) at the Medical School of São Paulo State University (FMB-UNESP). Sequencing quality was assessed by FastQC. Reads were aligned to the human genome (hg19, GRCh37) with BWA software, and SAM tools converted the alignment results to BAM format.[27] Next, the mapped reads underwent variant calling for SNP with GATK command line tools, including HaplotypeCaller, SelectVariants, and VariantFiltration programs with default parameters. After the calling step, the variants were annotated using the VEP[28] software. Coverage depth was a priori set at 100×. Variants had to have >10 reads of position depth (PD) and/or >6 reads of allele depth (AD) and/or an AD/PD ratio of >0.05 and/or a population frequency higher than 1% (popfreq_all >0.01) were included in the study. Finally, variants were compared using ABraOM, a web-based public database of Brazilian genomic variants.[29]

In silico Prediction of Non-Coding Genetic Variants

Several tools were used to predict potential functional effects of SNPs on non-coding binding sites, such as splice sites and binding sites for transcription factors, exonic splicing enhancers (ESEs), and microRNA (miRNA). The impact of each genetic variant was assessed using VarSome[30] an integrated search engine that allows access to several databases, forecasting tools, and publications on a single website. Variant pathogenicity was reported using an automatic variant classifier that evaluates each submitted variant according to guideline of the American College of Medical Genetics and Genomics (ACMG) and classifies it as either "pathogenic", "likely pathogenic", "likely benign", "benign" or "uncertain significance". Varsome predicts the pathogenicity of each variant through a DANN[31] score, a methodology for scoring deleterious annotations of genetic variants using neural networks that results in a number ranging from 0 to 1. Higher DANN scores represent greater variant deleteriousness.[31]

Next, we applied the Genomic Evolutionary Rate Profiling (GERP)[32,33] conservation score. This score is used to calculate the reduction of substitutions in a multi-species sequence alignment when compared to a neutral expectation. GERP scores >5.5 are strongly associated with a purifying selection. Mutations that occur at highly conserved sites in many species are assumed as harmful and therefore contribute to the genetic load within a species.

Finally, we used two tools, SNPinfo (FuncPred)[34] and RegulomeDB,[35] to track SNPs according to their functions. SNPinfo is a web server that helps researchers investigate SNPs in studies of genetic association and provide different pipelines for SNP selection, whereas RegulomeDB is an online composite database and prediction tool to annotate and prioritize potential regulatory variants from the human genome.[35] RegulomeDB divides the variants into six categories: category 1 variants are likely to affect binding and are linked to the expression of a gene target, category 2 variants are likely to affect binding, category 3 variants are less likely to affect binding, and category 4, 5, and 6 variants have minimal binding evidence.[35]

Statistical Analysis

The allelic and genotypic frequencies of the *PD-L1* polymorphisms found in NSCLC were calculated by Hardy Weinberg equilibrium ([1 _ (hC 2H)]/2N, where "h" stands for a heterozygous genotype, "H" for homozygous genotype and "N" for the number of samples). Associations between polymorphisms, PD-L1 protein expression, and the clinicopathological parameters of NSCLC patients were investigated by Chi-square test. The prognostic value of each polymorphism was assessed by a survival analysis using the Kaplan–Meier method with the Log rank test for statistical significance. In addition, Cox's proportional hazards regression models were used in a multivariate analysis to test the association between SNPs and PPS and OS. PPS was considered as the period from tumor progression until death or last follow-up. OS was defined as the time from curative surgery to death or last date known to be alive. The statistical software program IBM SPSS (version 22; Armonk, NY, USA) performed all analyses. Differences were considered statistically significant at $P<0.05$.

Results

Clinicopathologic Characteristics

Of the 70 patients included in the study, 33 presented with ADC (47.1%), 24 with SqCC (34.3%), and 13 with LCC (18.6%). The clinical characteristics by histologic types are summarized in Table 1. While SqCC cases were more frequent in males (81.8%), ADC cases were equally distributed between genders, and LCC cases were close to equal

Table 1 Demographic and Clinicopathological Characteristics of 70 NSCLC Patients

Characteristics	Histological Subtypes			*P*-value
	Adenocarcinoma (N=33)	Squamous Cell Carcinoma (N=24)	Large Cell Carcinoma (N=13)	
[a]Gender				
Male	10 (50.0%)	9 (81.8%)	5 (55.6%)	0.21
Female	10 (50.0%)	2 (18.2%)	4 (44.4%)	
[a]Age (years)				
≤ 63	11 (61.1%)	6 (54.5%)	6 (60.0%)	0.93
> 63	7 (38.9%)	5 (45.5%)	4 (40.0%)	
[a]Smoke Status				
Yes	8 (72.7%)	8 (88.9%)	3 (50.0%)	0.25
No	3 (27.3%)	1 (11.1%)	3 (50.0%)	
[a]Clinical Stage†				
I–II	8 (47.1%)	4 (40.0%)	3 (42.9%)	0.93
III–IV	9 (52.9%)	6 (60.0%)	4 (57.1%)	
[a]Treatment				
Chemotherapy				
Yes	7 (36.8%)	3 (27.3%)	4 (40.0%)	0.80
No	12 (63.2%)	8 (72.7%)	6 (60.0%)	
Radiotherapy				
Yes	1 (5.3%)	1 (9.1%)	0 (0.0%)	0.63
No	18 (94.7%)	10 (90.9%)	10 (100%)	
[a]Relapse				
Yes	9 (64.3%)	4 (50.0%)	2 (33.3%)	0.43
No	5 (35.7%)	4 (50.0%)	4 (66.7%)	
[a]Mcs+PD-L1 (median cell density)				
≤1.26	13 (54.2%)	10 (52.6%)	3 (30.0%)	0.40
>1.26	11 (45.8%)	9 (47.4%)	7 (70.0%)	
Follow-up (months) 66 (12–144)				
[a]Patients censored for survival analysis at last follow-up time	8 (53.3%)	4 (66.7%)	3 (42.9%)	-

Notes: [a]Some cases had missing follow-up information: gender (30); age (31); smoke status (44); clinical stage (36); adjuvant treatment QT/RT (30); relapse (42); Mcs+PD-L1 (17); survival (44). † 8th International Association for the Study of Lung Cancer [23].
Abbreviations: Mcs, malignant cells; PD-L1, programmed death-ligand 1.

distribution (male 55.6%, female 44.4%). All histological types were more frequent in patients aged 63 years or younger. 8 patients reported a history of tobacco smoking in the ADC group (72.7%) and in the SqCC group (88.9%), versus 3 patients in the LCC group (50.0%). All the histological subtypes included advanced stages of disease (9 cases in ADC, 6 cases in SqCC, and 4 cases in LCC). Most of the patients had not received either chemotherapy (12 cases to ADC, 8 cases to SqCC, and 6 cases to LCC) or radiation therapy (18 cases to ADC, 10 cases to SqCC, and LCC) as adjuvant treatment. Malignant cells expressed PD-L1 above the median in 7 LCC cases (70.0%), the most relevant expression compared to the other two histological subtypes. The median follow-up was 66 (12–144) months. None of the analysis revealed significant differences between histological types ($P>0.05$).

Allele and Genotype Distributions of PD-L1 Gene Polymorphisms

All NSCLC patients who underwent surgical resection were successfully genotyped for fifteen *PD-L1* SNPs: rs76805387T>C, rs4742098A>G, rs47946526A>G, rs10217310G>T, rs7864231G>A, rs41280725C>T, rs573692330A>G, rs1011769981G>A, rs41280723T>C, rs138135676T>C, rs4143815G>C, rs2297136G>A,

rs148242519G>A, rs41303227C>T, and rs7041009G>A. Supplementary Table 1 shows the SNP identification numbers, allele and genotype frequencies, and *P*-value for HWE. Of the 15 SNPs studied, 11 were found to be monomorphic, whereas 4 SNPs, namely rs4742098, rs4143815, rs2297136, and rs7041009, were polymorphic in NSCLC. Monomorphic SNPs were excluded from further analysis. All the polymorphic SNPs were found to be in equilibrium (P>0.05) for HWE. The allele frequency of our cohort was compared to different populations in the 1000 Genomes Project (Supplementary Table 2).

Correlation Between PD-L1 Gene Polymorphisms and Clinicopathological Characteristics in NSCLC

We performed stratified analyses on the associations between clinical characteristics and the four *PD-L1* polymorphisms with different genotypic distributions. Table 2 shows each SNP genotype frequency and their associated clinicopathological characteristics. Three of the four *PD-L1* gene polymorphisms (rs4742098, rs4143815, and rs7041009) were significantly associated with relapse

Table 2 Clinicopathological Characteristics of 70 NSCLC Patients Stratified by the *PD-L1* Polymorphisms rs4742098, rs4143815, rs2297136 and rs7041009

Characteristics	rs4742098 AA vs AG/GG			rs4143815 GG vs CG/CC			rs2297136 GG vs AG/AA			rs7041009 GG vs AG/AA		
	No. of Patients (N=70) (%)			No. of Patients (N=70) (%)			No. of Patients (N=70) (%)			No. of Patients (N=70) (%)		
	A/A	A/G+G/G	*P*	G/G	C/G+C/C	*P*	G/G	AG+A/A	*P*	G/G	A/G+A/A	*P*
Age (years), median												
≤ 63	13 (56.5%)	10 (62.5%)	0.75	11 (57.9%)	12 (60.0%)	1.00	4 (57.1%)	19 (59.4%)	1.00	16 (76.2%)	7 (38.9%)	**0.02**
>63	10 (43.5%)	6 (37.5%)		8 (42.1%)	8 (40.0%)		3 (42.9%)	13 (40.6%)		5 (23.8%)	11 (61.1%)	
Gender, n (%)												
Male	13 (56.5%)	11 (64.7%)	0.74	12 (63.2%)	12 (57.1%)	0.75	4 (57.1%)	20 (60.6%)	1.00	12 (57.1%)	12 (63.2%)	0.75
Female	10 (43.5%)	6 (35.3%)		7 (36.8%)	9 (42.9%)		3 (42.9%)	13 (39.4%)		9 (42.9%)	7 (36.8%)	
Smoke status												
Yes	10 (71.4%)	9 (75.0%)	1.00	9 (75.0%)	10 (71.4%)	1.00	4 (66.7%)	15 (75.0%)	1.00	9 (60.0%)	10 (90.9%)	0.17
No	4 (28.6%)	3 (25.0%)		3 (25.0%)	4 (28.6%)		2 (33.3%)	5 (25.0%)		6 (40.0%)	1 (9.1%)	
Histology												
ADC	18 (43.9%)	15 (51.7%)	0.61	15 (41.7%)	18 (52.9%)	0.40	7 (53.8%)	26 (45.6%)	0.86	15 (41.7%)	18 (55.9%)	0.34
SqCC	16 (39.0%)	8 (27.6%)		15 (41.7%)	9 (26.5%)		4 (30.8%)	20 (35.1%)		12 (33.3%)	12 (35.3%)	
LCC	7 (17.1%)	6 (20.7%)		6 (16.7%)	7 (20.6%)		2 (15.4%)	11 (16.3%)		9 (25.0%)	4 (11.8%)	
Clinical stage†												
I–II	8 (40.0%)	7 (50.0%)	0.72	7 (41.2%)	8 (47.1%)	1.00	1 (14.3%)	14 (51.9%)	0.10	8 (44.4%)	7 (43.8%)	1.00
III–IV	12 (60.0%)	7 (50.0%)		10 (58.8%)	9 (52.9%)		6 (85.7%)	13 (48.1%)		10 (55.6%)	9 (56.3%)	
Mcs+PD-L1 (median)												
≤1.26	15 (57.7%)	11 (42.3%)	0.78	14 (53.8%)	12 (46.2%)	1.00	6 (23.1%)	20 (76.9%)	0.74	11 (42.3%)	15 (57.7%)	0.27
>1.26	14 (51.9%)	13 (48.1%)		14 (51.9%)	13 (48.1%)		5 (18.5%)	22 (81.5%)		16 (59.3%)	11 (40.7%)	
Chemotherapy												
Yes	7 (30.4%)	7 (41.2%)	0.52	5 (26.3%)	9 (42.9%)	0.33	3 (42.9%)	11 (33.3%)	0.67	10 (47.6%)	4 (21.1%)	0.10
No	16 (69.6%)	10 (58.8%)		14 (73.7%)	12 (57.1%)		4 (57.1%)	22 (66.7%)		11 (52.4%)	15 (78.9%)	
Radiotherapy												
Yes	1 (4.3%)	1 (5.9%)	1.00	1 (5.3%)	1 (4.8%)	1.00	1 (14.3%)	1 (3.00%)	0.32	0 (0.0%)	2 (10.5%)	0.21
No	22 (95.7%)	16 (94.1%)		18 (94.7%)	20 (95.2%)		6 (85.7%)	32 (97.0%)		21 (100.0%)	17 (89.5%)	
Relapse												
Yes	11 (78.6%)	4 (28.6%)	**0.01**	9 (75.0%)	6 (37.5%)	**0.05**	5 (71.4%)	10 (47.6%)	0.39	5 (33.3%)	10 (76.9%)	**0.02**
No	3 (21.4%)	10 (71.4%)		3 (25.0%)	10 (62.5%)		2 (28.6%)	11 (52.4%)		10 (66.7%)	3 (23.1%)	
Status												
Live	5 (38.5%)	8 (53.3%)	0.47	5 (45.5%)	8 (47.1%)	1.00	3 (50.0%)	10 (45.5%)	1.00	11 (78.6%)	2 (14.3%)	**0.00**
Dead	8 (61.5%)	7 (46.7%)		6 (54.5%)	9 (52.9%)		3 (50.0%)	12 (54.5%)		3 (21.4%)	12 (85.7%)	

Notes: Bolded values are statically significant. †8th International Association for the Study of Lung Cancer [23].
Abbreviations: ADC, adenocarcinoma; SqCC, squamous cell carcinoma; LCC, large cell carcinoma; Mcs, malignant cells; PD-L1, programmed death-ligand 1.

(*P*=0.01; *P*=0.05; *P*=0.02, respectively). For the rs4742098 variant, carriers of the G allele (AG or GG genotypes) were less likely to relapse (*P*=0.01). Similarly, for rs4143815, carriers of the alternative C allele (CG or CC genotypes) were also less likely to relapse (*P*=0.05). In rs7041009, however, carriers of the alternative allele A (AG or GG genotypes) were more likely to relapse (*P*=0.02). Moreover, GG genotype (reference) of rs7041009 showed a significant correlation with age, being more prevalent among younger patients (16 patients, or 69.6%), and status, being more prevalent among patients who were alive (11 patients, or 84.6%), compared to carriers of the A allele (*P*=0.02 and *P*<0.01, respectively). No statistical significance was observed in the association between rs2297136 genotypes and clinicopathological variants.

Correlation Between PD-L1 Gene Polymorphisms and PD-L1 Protein Expression

The correlation between PD-L1 protein expression and *PD-L1* gene polymorphisms are shown in Table 2. There were no statistically significant associations between PD-L1 protein expression in malignant cells and *PD-L1* gene polymorphisms. In our cohort, the four *PD-L1* gene polymorphisms were in non-coding regions and, apparently, cause no interference in PD-L1 protein expression in NSCLC malignant cells. However, when we correlated PD-L1 protein expression with histological subtype, we observed that the expression in malignant cells was above the median in 70% of patients with LLC, in contrast with 45.8% and 47.4% of patients with ADC and SqCC, respectively.

Associations Between PD-L1 Gene Polymorphisms and Survival Outcomes

Our first statistical test examined the individual effect of patients' characteristics to estimate statistical differences in survival using the Kaplan–Meier method (Table 3). Patients younger than 63 years showed increased OS, 111.62 vs 66.54 months in older patients (*P*=0.05). Choice of treatment was also an independent factor in diagnosis, with patients who did not receive radiotherapy presenting a better survival rate when compared with those who were treated with radiotherapy, 94.43 vs 12.00 months, respectively (*P*<0.01). Patients who presented disease recurrence had lower survival rates and poorer prognostic when compared with those who did not relapse, 48.23 vs 123.10 months, respectively (*P*<0.01).

Table 3 A Survival Analysis Conducted by the Kaplan–Meier Method Showing the Difference in the Means of the Log Rank Test According to the Optimal Upper and Lower Binary Cut-off Limits of Different Variables

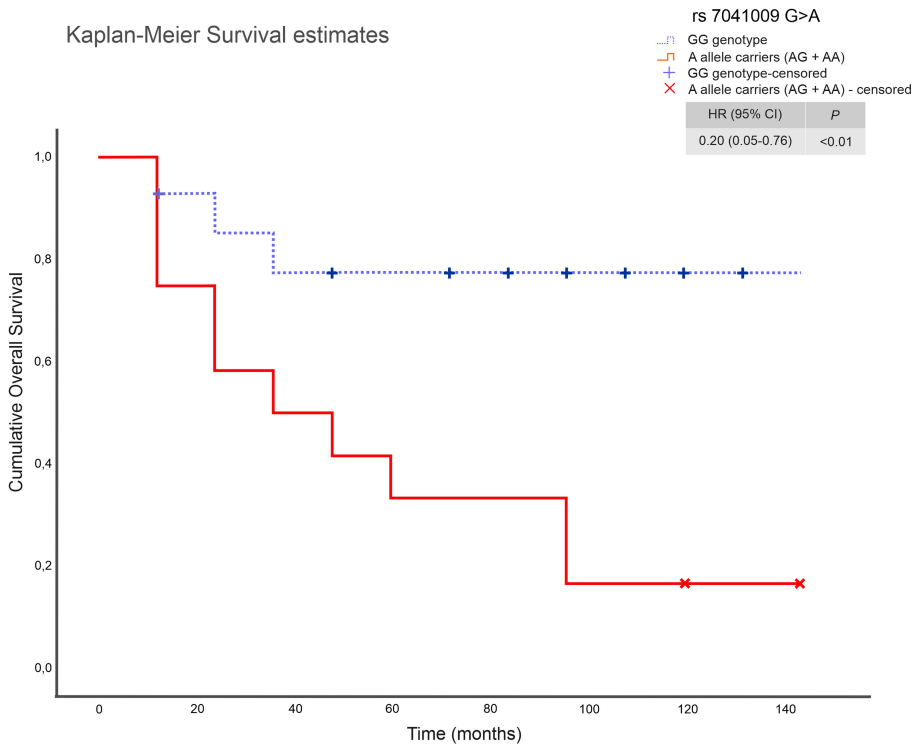
Variables	Overall Survival (Months)		Chi-Square (Log Rank)	*P*-value
	Mean	Standard Error		
Age				
≤63	111.62	13.92	3.88	0.05
>63	66.54	15.99		
Treatment				
Radiotherapy				
Yes	12.00	0.00	11.46	**<0.01**
No	94.43	11.31		
Relapse				
Yes	48.23	9.84	11.01	**<0.01**
No	123.10	13.48		
***PD-L1* polymorphisms**				
rs7041009 (GG vs AG/AA genotype)				
GG	116.93	13.81	7.49	**<0.01**
AG/AA	59.00	13.67		

Note: Bolded values are statically significant.
Abbreviation: PD-L1, programmed death-ligand 1.

Moreover, differences in the genotypes of *PD-L1* polymorphisms seemed to also impact the prognosis of NSCLC patients. *PD-L1* rs7041009, for instance, led to a statistically significant difference in OS (Figure 1), with carriers of the A allele of rs7041009 having lower OS than carriers of the GG genotype (reference), 59.00 vs 116.93 months, respectively (*P*<0.01).

Next, using a univariate Cox Regression analysis, we were able to associate the following variables with a lower risk of death: the absence of radiotherapy treatment, relapse, and GG genotype of *PDL1* rs7041009 (Table 4). However, after feeding these variables into a multivariate analysis, only the absence of radiotherapy treatment and relapse were considered to be independent factors for OS (HR 9.82, *P*=0.02; HR 6.15, *P*=0.04, respectively).

Then, we introduced the *PD-L1* polymorphisms into the Cox model, controlling for radiotherapy treatment and tumor relapse. Of the four SNPs, only rs7041009 was identified as a co-dependent factor associated with radiotherapy and relapse. We thus inferred that patients with NSCLC who



rs 7041009 G>A	Survival estimates (months)											
	12	24	36	48	60	72	84	96	108	120	132	144
GG genotype	60.0%	33.3%	50.0%	50.0%	0.0%	100.0%	100.0%	33.3%	100.0%	50.0%	100.0%	50.0%
AG genotype	40.0%	33.3%	50.0%	50.0%	100.0%	0.0%	0.0%	67.7%	0.0%	50.0%	0.0%	0.0%
AA genotype	0.0%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	50.0%

Figure 1 Kaplan–Meier survival curve for *PD-L1* rs7041009 G>A. A allele carriers (AG+AA) presented worse prognosis and a lower survival rate when compared to GG genotyped patients ($P<0.01$).

carried the A allele (AG/AA) presented a higher risk of relapse in the presence of radiotherapy, resulting in a poorer prognosis and decreased survival rates than patients who carried the rs7041009 GG genotype. In relapsed patients, we observed that the *PD-L1* polymorphisms rs7041009 and rs4742098 might have an impact on PPS (Figure 2). Patients with the rs7041009 GG genotype had a higher PPS than those with the alternative A allele of rs7041009 (AG/AA), 110.98 vs 56.18 months, respectively ($P<0.01$); whereas, patients who carried the reference rs4742098 AA genotype had lower PPS than those who carried the alternative G allele of rs4742098 (AG/GG), 56.00 vs 115.71 months, respectively ($P=0.02$).

In silico Prediction of PD-L1 Gene Polymorphisms

The in silico analysis predicted the *PD-L1* variants rs4742098 (c.*2635A>G), rs4143815 (c.*395G>C), rs2297136 (c.*93G>A), and rs7041009 (c.682+122G>A) to be benign (Table 5). Not only was the DANN score low for all four variants (0.8226, 0.6475, 0.7056, and 0.5428, respectively), but their GERP score was also lower than 5.5 (−1.74, 2.38, 4.4, and 1.81, respectively), indicating that these variants are found in non-conserved positions and are unlikely to be harmful.

SNPinfo predicted miRNA-binding function to be affected by two of these variants, namely rs2297136 and rs4143815. Rs2297136 was predicted to affect the binding function of hsa-miR-324-5p and hsa-miR-632, whereas rs4143815 was found to correlate with hsa-miR-1252, hsa-miR-1253, hsa-miR-539, hsa-miR-548, and hsa-miR-570 (Table 6). RegulomeDB was then used to complement the SNP analysis. Three of the four SNPs, rs4742098, rs4143815, and rs2297136, had a RegulomeDB score of 5, whereas rs7041009 had a score of 6, meaning that all four variants show minimal binding evidence (Table 5).

Discussion

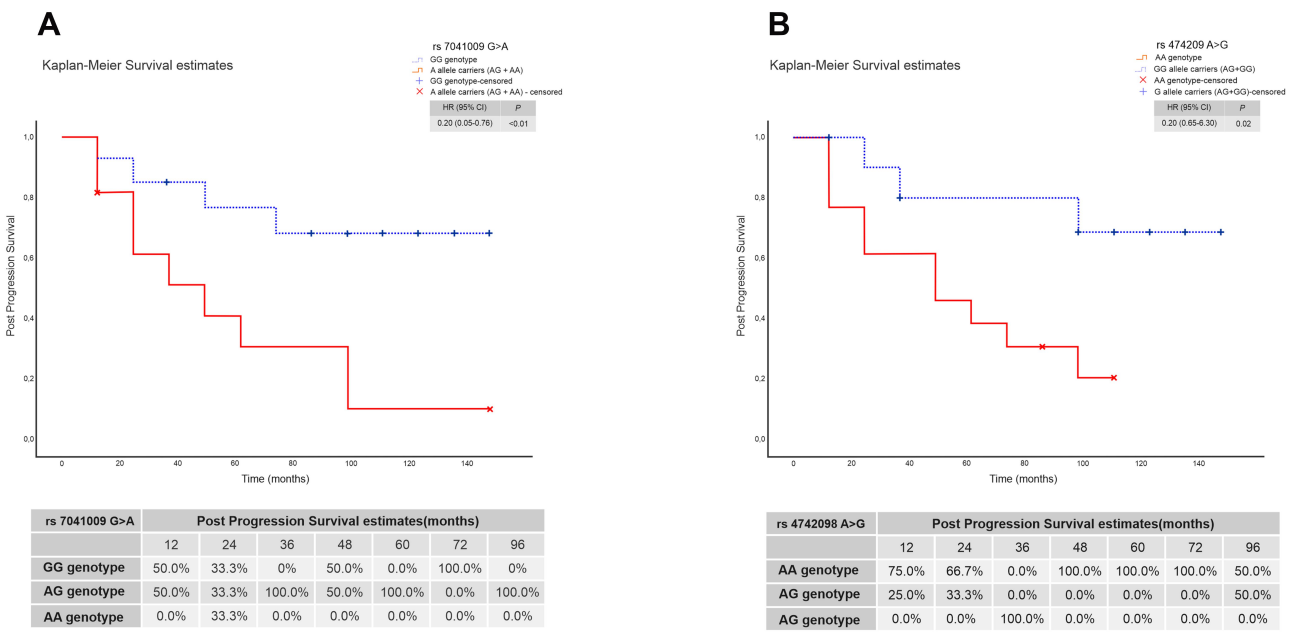
Lung cancer has a high mortality rate and lacks suitable markers for early diagnosis and prognosis. Thus, it is essential to detect the best potential biomarker out of the several genetic and protein markers. Fortunately for

Table 4 Variables Associated with Overall Survival (OS) in 70 Patients Diagnosed with NSCLC. Univariate and Multivariate Analyses Employing a Cox Proportional Hazards Model

Clinicopathological Characteristics	Univariate Analysis			Multivariate Analysis	
	HR (95% CI)	HR	*P* value	HR[c] (95% CI)	*P*-value
Age (yrs): ≤63 vs >63	0.33 (0.10–1.10)	−1.10	0.07		
Gender					
Male	1.83 (0.54–6.14)	0.60	0.32		
Smoke status					
Yes	1.33 (0.33–5.36)	0.28	0.68		
Histology					
ADC	1.25 (0.32–4.86)	0.22	0.74		
SqCC	1.67 (0.33–8.37)	0.51	0.53		
LCC (reference)	-	-	0.82		
Clinical stage†					
I–II	0.41 (0.11–1.52)	−0.87	0.18		
III–IV (reference)	-	-	-		
Therapy					
Chemotherapy	0.36 (0.10–1.34)	−1.00	0.13		
Yes					
Relapse	8.47 (1.81–39.62)	2.13	**0.00**	6.15 (1.03–36.44)	**0.04**
Yes					
Immune checkpoint					
PD-L1+/mm^2	1.00 (0.95–1.05)	0.00	0.94		
polymorphisms in PD-L1					
rs4742098					
AA genotype	2.02 (0.65–6.30)	0.70	0.22		
G allele carrier (reference)	-	-	-		
rs4143815					
GG genotype	1.58 (0.52–4.81)	0.46	0.41		
C allele carrier (reference)	-	-	-		
rs2297136					
GG genotype	1.24 (0.34–4.55)	0.22	0.73		
A allele carrier (reference)	-	-	-		
rs7041009					
GG genotype	0.20 (0.05–0.76)	−1.56	**0.01**	0.57 (0.12–2.70)	0.48
A allele carrier (reference)	-	-	-		

Notes: Bolded values are statically significant. †8th International Association for the Study of Lung Cancer [23].
Abbreviations: ADC, adenocarcinoma; SqCC, squamous cell carcinoma; LCC, large cell carcinoma; HR, hazard ratio (β coefficient); CI, confidence interval; PD-L1, programmed death-ligand 1.

patients, the translational impact of such findings is rapidly increasing, and the stimulation of immune response by ICIs has emerged as a dramatic paradigm shift in the treatment of advanced tumors, mainly NSCLC.[19,36] PD-1/PD-L1 monoclonal antibodies have shown potential efficacy in advanced squamous-cell and non-squamous NSCLC.[37,38] However, despite the remarkable success achieved by immunotherapy so far, its effectiveness still seems to vary among cancer patients.[19] The expression of PD-L1 on tumor cells remains the only recognized predictive factor for immunotherapy response in NSCLC patients; however, patients without PD-L1 expression on



rs 7041009 G>A	Post Progression Survival estimates(months)						
	12	24	36	48	60	72	96
GG genotype	50.0%	33.3%	0%	50.0%	0.0%	100.0%	0%
AG genotype	50.0%	33.3%	100.0%	50.0%	100.0%	0.0%	100.0%
AA genotype	0.0%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%

rs 4742098 A>G	Post Progression Survival estimates(months)						
	12	24	36	48	60	72	96
AA genotype	75.0%	66.7%	0.0%	100.0%	100.0%	100.0%	50.0%
AG genotype	25.0%	33.3%	0.0%	0.0%	0.0%	0.0%	50.0%
AG genotype	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%

Figure 2 Kaplan–Meier survival curves estimating post-progression survival (PPS) in NSCLC patients according to *PD-L1* polymorphisms. **(A)** Kaplan–Meier survival curve for rs7041009 G>A. A allele carriers (AG+AA) presented worse prognosis and a lower PPS rate when compared GG genotyped patients ($P<0.01$); **(B)** Kaplan–Meier survival curve for rs4742098 A>G. G allele carriers (AG+GG) had a higher PPS rate and better prognosis when compared AA genotyped patients ($P=0.02$).

tumor cells may also respond to immunotherapy.[39–43] Based on these findings, the present study inferred that *PD-L1* non-coding genetic variants could help predict the prognosis of patients with NSCLC and impact disease recurrence and OS.

In our cohort of 70 NSCLC specimens, we evaluated PD-L1 protein expression in malignant cells by PD-L1 multiplex immunofluorescence (mIF) assays, using the Cell Signaling E1L3N clone and image analysis, and investigated *PD-L1* polymorphisms by NGS sequencing. This method resulted in the detection of high PD-L1 expression in LCC malignant cells when compared to other histological subtypes, suggesting that LCC patients may benefit from ICIs. As described by Shimoji et al,[44] PD-L1 expression using the Cell Signaling E1L3N clone was significantly correlated with a consistent vimentin expression, increased Ki-67 labeling index and poor prognosis in ADC but not in SqCC. Other studies have reported that PD-L1 was detected at significantly higher frequencies in SqCC than in ADC of the lung.[44,45] Cha et al[46] found that PD-L1 expression using the SP142 clone was significantly associated with an ADC solid subtype histology, p53 aberrant expression, and poor prognosis.

We also assessed *PD-L1* polymorphisms. Of the fifteen genetic variants genotyped, eleven were monomorphic. The four potential variants (rs4742098, rs4143815, rs2297136, and rs7041009) present in the non-coding region were correlated with the clinicopathological characteristics of the NSCLC patients and were submitted to an in silico analysis investigating their functional role. The MAF of the rs4742098, rs4143815, and rs7041009 polymorphisms was consistent with the genotype frequency among European and Mixed Americans populations present in The 1000 Genomes Project, whereas the G allele in

Table 5 List of the Selected Non-Coding SNP and the Tools Used to Study Them

dbSNP ID	Region	FuncPred	RegulomeDB score	VarSome (DANN Score)
rs4742098	3′UTR	–	5	Benign (0.8226)
rs4143815	3′UTR	miRNA-binding function	5	Benign (0.6475)
rs2297136	3′UTR	miRNA-binding function	5	Benign (0.7056)
rs7041009	Intron	–	6	Benign (0.5428)

Note: SNPs that do not affect function.
Abbreviations: SNPs, single-nucleotide polymorphisms; dbSNP, SNP database; FuncPred, SNP function prediction; RegulomeDB score, RegulomeDB variant classification; VarSome, The human genomics community; DANN score, annotation of genetic variants using neural networks.

Table 6 List of the 3′ UTR SNPs Analyzed in FuncPred and Their miRNA Motif

dbSNP ID	Allele	miRNA Motif
rs4143815	C C C C G	*hsa-miR-1252* *hsa-miR-1253* *hsa-miR-539* *hsa-miR-548* *hsa-miR-570*
rs2297136	A/G A/G	*hsa-miR-324-5p* *hsa-miR-632*

Abbreviations: FuncPred, SNP function prediction; dbSNP ID, SNP database identification.

the rs2297136 polymorphism was the main allele in our cohort to show racial differences.

When assessing patient prognosis, three of the four *PD-L1* variants rs4742098, rs4143815, and rs7041009 were significantly associated with disease recurrence. Carriers of the G allele (individuals with the AG or GG genotypes) of rs4742098 were less likely to relapse compared to carriers of the homozygous AA genotype. Similar findings were published by Du and colleagues,[47] who reported that the AG genotype differed from the AA genotype in terms of risk of NSCLC recurrence. In the case of rs4143815G, patients with the alternative C allele were less likely to relapse in our study, in agreement with Nomizo et al's report.[48] In their study, the authors even suggested that this polymorphism might be a biomarker for nivolumab efficacy.[48] Finally, in our study, for rs7041009G>A, carriers of the alternative G allele were more likely to exhibit relapse. The rs7041009 GG genotype also showed a significant correlation with age, being more present in younger patients, and with status, being more present in patients who are alive, compared to carriers of the A allele. Rs7041009 (c.682+122G>A) is located at position 2377 in intron 4 of the PD-L1 gene. However, little is known about the exact function of this genetic variation, except that it is located near the transcription factor binding site.

Our cohort showed a significant association between these PD-L1 polymorphisms and OS in NSCLC. 18 Patients presenting the GG genotype of rs7041009 benefited from longer OS. In addition, our findings indicated that both the rs7041009G and rs4742098G polymorphisms were significantly related to a longer PPS. The clinical impacts of *PD-L1* variants had also been investigated by previous studies. Zhao et al[49] suggested that patients with the GG genotype of another *PD-L1* polymorphism (rs822336) had worse disease-free survival and OS in a Chinese patient population.

Further contribution was provided in that respect by Lee et al[50] who demonstrated that rs4143815 and rs2297136 were significantly associated with clinical outcomes after chemotherapy. Of 379 patients with NSCLC treated with first-line paclitaxel–cisplatin chemotherapy, those carrying the rs4143815 G allele responded better to chemotherapy and had gains in overall survival. In our study, however, the polymorphism rs2297136 showed no significant association with clinical outcome, a finding that was corroborated by Zhao et al.[49] This difference might be explained by the heterogeneity of patients enrolled in each study, and further research is needed to settle this question.

In our study, we were unable to find any statistical significance between the rs4742098, rs4143815, rs2297136, and rs7041009 genotypes and PD-L1 protein expression in malignant cells. However, recent data have helped to shed light on the impact of *PD-L1* genetics on PD-L1 expression, though the existing results remain controversial. Recently, Tao and colleagues[51] showed that rs4143815 and rs10815225 in the *PD-L1* gene contributed to PD-L1 overexpression in gastric cancer. So far, Lee et al[18] have conducted the largest study on *PD-L1* polymorphisms and PD-L1 expression in NSCLC. The authors showed that rs822336C, rs822337A and rs4143815G were associated with worse OS in NSCLC patients but found no significant correlation between PD-L1 expression and the genotypes of these polymorphisms. Krawczyk et al[52] demonstrated that carriers of the rs822335 CC genotype were predisposed to higher expression of the PD-L1 protein in NSCLC tumor cells, whereas rs822336 had no effect on PD-L1 expression in these cells. Their results are consistent with our findings, but future research is needed to clarify remaining confounders.

In our study, using in silico approaches, we report that rs4742098, rs4143815, rs2297136, and rs7041009 can be considered benign variants. However, little is known about the effect that mutations in conserved non-coding regions might have on fitness and how many of them are present in the human genome as deleterious polymorphisms. Moreover, rs2297136 was predicted to affect the miRNA-binding function of hsa-miR-324-5p and hsa-miR-632, whereas the variant rs4143815 was found to correlate with hsa-miR-1252, hsa-miR-1253, hsa-miR-539, hsa-miR-548, and hsa-miR-570. In this context, others have

reported that variants in the 3′UTR region of the *PD-L1* gene can affect the interaction of miRNAs, possibly resulting in *PD-L1* underexpression.[53] Therefore, additional studies are necessary to validate our findings. However, there are limitations to our analysis. We did not perform a case–control approach and our cohort comprises a relatively small sample size. Nonetheless, to our knowledge, our research on the effect of the rs7041009 polymorphism of *PD-L1* gene on NSCLC patients is unique.

We believe this study is also the first to evaluate variants in the non-coding region of the *PD-L1* in Brazilian patients with NSCLC, since most studies of *PD-L1* polymorphisms have been conducted in Asian patients. Thus, we consider this exploratory study as a pioneer in the understanding of *PD-L1* polymorphisms in a genetic admixed population.

Acknowledgments

We appreciate all subjects who participated in this study and the Illumina members for assistance with the initial runs.

Funding

This study was supported by the São Paulo Research Foundation, FAPESP (grant numbers 2013/14277-4 and 2018/20403-6).

Disclosure

The authors report no conflicts of interest related to this work.

References

1. Shankar A, Saini D, Dubey A, et al. Feasibility of lung cancer screening in developing countries: challenges, opportunities and way forward. *Transl Lung Cancer Res*. 2019;8(Suppl S1):S106–S121. doi:10.21037/tlcr.2019.03.03
2. Berghmans T, Durieux V, Hendriks LEL, Dingemans A-M. Immunotherapy: from Advanced NSCLC to Early Stages, an Evolving Concept. *Front Med*. 2020;7:90. doi:10.3389/fmed.2020.00090
3. Bylicki O, Paleiron N, Rousseau-Bussac G, Chouaïd C. New PDL1 inhibitors for non-small cell lung cancer: focus on pembrolizumab. *Onco Targets Ther*. 2018;11:4051–4064. doi:10.2147/OTT.S154606
4. Yeo M-K, Choi S-Y, Seong I-O, Suh K-S, Kim JM, Kim K-H. Association of PD-L1 expression and PD-L1 gene polymorphism with poor prognosis in lung adenocarcinoma and squamous cell carcinoma. *Hum Pathol*. 2017;68:103–111. doi:10.1016/j.humpath.2017.08.016
5. Francisco LM, Salinas VH, Brown KE, et al. PD-L1 regulates the development, maintenance, and function of induced regulatory T cells. *J Exp Med*. 2009;206(13):3015–3029. doi:10.1084/jem.20090847
6. Lantuejoul S, Damotte D, Hofman V, Adam J. Programmed death ligand 1 immunohistochemistry in non-small cell lung carcinoma. *J Thorac Dis*. 2019;11(Suppl S1):S89–S101. doi:10.21037/jtd.2018.12.103
7. Aggarwal C, Abreu DR, Felip E, et al. Prevalence of PD-L1 expression in patients with non-small cell lung cancer screened for enrollment in KEYNOTE-001, −010, and −024. *Ann Oncol*. 2016;27(6):359–378. doi:10.1093/annonc/mdw378.14
8. Yu H, Boyle TA, Zhou C, Rimm DL, Hirsch FR. PD-L1 Expression in Lung Cancer. *J Thorac Oncol*. 2016;11(7):964–975. doi:10.1016/j.jtho.2016.04.014
9. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1–Positive Non–Small-Cell Lung Cancer. *N Engl J Med*. 2016;375(19):1823–1833. doi:10.1056/NEJMoa1606774
10. Marwitz S, Scheufele S, Perner S, Reck M, Ammerpohl O, Goldmann T. Epigenetic modifications of the immune-checkpoint genes CTLA4 and PDCD1 in non-small cell lung cancer results in increased expression. *Clin Epigenetics*. 2017;9(1):51. doi:10.1186/s13148-017-0354-2
11. Chen L, Gibbons DL, Goswami S, et al. Metastasis is regulated via microRNA-200/ZEB1 axis control of tumour cell PD-L1 expression and intratumoral immunosuppression. *Nat Commun*. 2014;5(1):5241. doi:10.1038/ncomms6241
12. Ma Y, Adjemian S, Mattarollo SR, et al. Anticancer chemotherapy-induced intratumoral recruitment and differentiation of antigen-presenting cells. *Immunity*. 2013;38(4):729–741. doi:10.1016/j.immuni.2013.03.003
13. Mazzaschi G, Madeddu D, Falco A, et al. Low PD-1 Expression in Cytotoxic CD8 + Tumor-Infiltrating Lymphocytes Confers an Immune-Privileged Tissue Microenvironment in NSCLC with a Prognostic and Predictive Value. *Clin Cancer Res*. 2018;24(2):407–419. doi:10.1158/1078-0432.CCR-17-2156
14. de Vooght KMK, van Wijk R, van Solinge WW. Management of Gene Promoter Mutations in Molecular Diagnostics. *Clin Chem*. 2009;55(4):698–708. doi:10.1373/clinchem.2008.120931
15. Cuykendall TN, Rubin MA, Khurana E. Non-coding genetic variation in cancer. *Curr Opin Syst Biol*. 2017;1:9–15. doi:10.1016/j.coisb.2016.12.017
16. Amlie-Wolf A, Tang M, Way J, et al. Inferring the Molecular Mechanisms of Noncoding Alzheimer's Disease-Associated Genetic Variants. *J Alzheimers Dis*. 2019;72(1):301–318. doi:10.3233/JAD-190568
17. Hashemi M, Karami S, Sarabandi S, et al. Association between PD-1 and PD-L1 Polymorphisms and the Risk of Cancer: A Meta-Analysis of Case-Control Studies. *Cancers*. 2019;11(8):1150. doi:10.3390/cancers11081150
18. Lee SY, Jung DK, Choi JE, et al. Functional polymorphisms in PD-L1 gene are associated with the prognosis of patients with early stage non-small cell lung cancer. *Gene*. 2017;599:28–35. doi:10.1016/j.gene.2016.11.007
19. Fabrizio FP, Trombetta D, Rossi A, Sparaneo AA, Castellana S, Muscarella LA. Gene code CD274/PD-L1: from molecular basis toward cancer immunotherapy. *Ther Adv Med Oncol*. 2018;10:1758835918815598. doi:10.1177/1758835918815598
20. Araujo LH, Baldotto CA, Castro JGD, et al. Lung cancer in Brazil. *J Bras Pneumol*. 2018;44(1):55–64. doi:10.1590/s1806-37562017000000135
21. de Melo AC, de Sá VK, Sternberg C, et al. Mutational Profile and New IASLC/ATS/ERS Classification Provide Additional Prognostic Information about Lung Adenocarcinoma: A Study of 125 Patients from Brazil. *Oncology*. 2015;89(3):175–186. doi:10.1159/000376552
22. de Sá VK, Coelho JC, Capelozzi VL, de Azevedo SJ. Lung cancer in Brazil: epidemiology and treatment challenges.. *Lung Cancer*. 2016;7:141–148. doi:10.2147/LCTT.S93604

23. Goldstraw P, Chansky K, Crowley J, et al. The IASLC Lung Cancer Staging Project: proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol*. 2016;11(1):39–51. doi:10.1016/j.jtho.2015.09.009
24. Parra ER, Uraoka N, Jiang M, et al. Validation of multiplex immunofluorescence panels using multispectral microscopy for immune-profiling of formalin-fixed and paraffin-embedded human tumor tissues. *Sci Rep*. 2017;7(1):13380. doi:10.1038/s41598-017-13942-8
25. Parra ER, Jiang M, Machado-Rugolo J, et al. Variants in Epithelial-Mesenchymal Transition and Immune Checkpoint Genes Are Associated With Immune Cell Profiles and Predict Survival in Non–Small Cell Lung Cancer. *Arch Pathol Lab Med*. 2020;144(10):1234–1244. doi:10.5858/arpa.2019-0419-OA
26. Gorris MAJ, Halilovic A, Rabold K, et al. Eight-Color Multiplex Immunohistochemistry for Simultaneous Detection of Multiple Immune Checkpoint Molecules within the Tumor Microenvironment. *J Immunol*. 2018;200(1):347–354. doi:10.4049/jimmunol.1701262
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760. doi:10.1093/bioinformatics/btp324
28. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor.. *Genome Biol*. 2016;17(1):122. doi:10.1186/s13059-016-0974-4
29. Naslavsky MS, Yamamoto GL, de Almeida TF, et al. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. *Hum Mutat*. 2017;38(7):751–763. doi:10.1002/humu.23220
30. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. *Bioinformatics*. 2019;35(11):1978–1980. doi:10.1093/bioinformatics/bty897
31. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*. 2015;31(5):761–763. doi:10.1093/bioinformatics/btu703
32. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol*. 2010;6(12):e1001025. doi:10.1371/journal.pcbi.1001025
33. Cooper GM, et al. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res*. 2005;15(7):901–913. doi:10.1101/gr.3577405
34. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res*. 2009;37(suppl_2):W600–W605. doi:10.1093/nar/gkp290
35. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790–1797. doi:10.1101/gr.137323.112
36. Nishino M, Ramaiya NH, Hatabu H, Hodi FS. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol*. 2017;14(11):655–668. doi:10.1038/nrclinonc.2017.88
37. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer. *N Engl J Med*. 2015;373(17):1627–1639. doi:10.1056/NEJMoa1507643
38. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non–Small-Cell Lung Cancer. *N Engl J Med*. 2015;373(2):123–135. doi:10.1056/NEJMoa1504627
39. Yi C, He Y, Xia H, Zhang H, Zhang P. <p>Review and perspective on adjuvant and neoadjuvant immunotherapies in NSCLC. *Onco Targets Ther*. 2019;12:7329–7336. doi:10.2147/OTT.S218321
40. Horn L, Spigel DR, Vokes EE, et al. Nivolumab Versus Docetaxel in Previously Treated Patients With Advanced Non–Small-Cell Lung Cancer: two-Year Outcomes From Two Randomized, Open-Label, Phase III Trials (CheckMate 017 and CheckMate 057). *J Clin Oncol*. 2017;35(35):3924–3933. doi:10.1200/JCO.2017.74.3062
41. Antonia SJ, Villegas A, Daniel D, et al. Durvalumab after Chemoradiotherapy in Stage III Non–Small-Cell Lung Cancer. *N Engl J Med*. 2017;377(20):1919–1929. doi:10.1056/NEJMoa1709937
42. Fehrenbacher L, von Pawel J, Park K, et al. Updated Efficacy Analysis Including Secondary Population Results for OAK: A Randomized Phase III Study of Atezolizumab versus Docetaxel in Patients with Previously Treated Advanced Non–Small Cell Lung Cancer. *J Thorac Oncol*. 2018;13(8):1156–1170. doi:10.1016/j.jtho.2018.04.039
43. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al. Pembrolizumab plus Chemotherapy in Metastatic Non–Small-Cell Lung Cancer. *N Engl J Med*. 2018;378(22):2078–2092. doi:10.1056/NEJMoa1801005
44. Shimoji M, Shimizu S, Sato K, et al. Clinical and pathologic features of lung cancer expressing programmed cell death ligand 1 (PD-L1). *Lung Cancer*. 2016;98:69–75. doi:10.1016/j.lungcan.2016.04.021
45. Takada K, Toyokawa G, Okamoto T, et al. A Comprehensive Analysis of Programmed Cell Death Ligand-1 Expression With the Clone SP142 Antibody in Non–Small-Cell Lung Cancer Patients. *Clin Lung Cancer*. 2017;18(5):572–582. doi:10.1016/j.cllc.2017.02.004
46. Cha YJ, Kim HR, Lee CY, Cho BC, Shim HS. Clinicopathological and prognostic significance of programmed cell death ligand-1 expression in lung adenocarcinoma and its relationship with p53 status. *Lung Cancer*. 2016;97:73–80. doi:10.1016/j.lungcan.2016.05.001
47. Du W, Zhu J, Chen Y, et al. Variant SNPs at the microRNA complementary site in the B7-H1 3′-untranslated region increase the risk of non-small cell lung cancer. *Mol Med Rep*. 2017;16(3):2682–2690. doi:10.3892/mmr.2017.6902
48. Nomizo T, Ozasa H, Tsuji T, et al. Clinical Impact of Single Nucleotide Polymorphism in PD-L1 on Response to Nivolumab for Advanced Non-Small-Cell Lung Cancer Patients. *Sci Rep*. 2017;7(1):45124. doi:10.1038/srep45124
49. Zhao M, Zhang J, Chen S, Wang Y, Tian Q. <p>Influence of Programmed Death Ligand-1-Gene Polymorphism rs822336 on the Prognosis and Safety of Postoperative Patients with NSCLC Who Received Platinum-Based Adjuvant Chemotherapy. *Cancer Manag Res*. 2020;12:6755–6766. doi:10.2147/CMAR.S255072
50. Lee SY, Jung DK, Choi JE, et al. PD-L1 polymorphism can predict clinical outcomes of non-small cell lung cancer patients treated with first-line paclitaxel-cisplatin chemotherapy. *Sci Rep*. 2016;6(1):25952. doi:10.1038/srep25952
51. Tao L-H, Zhou X-R, Li F-C, et al. A polymorphism in the promoter region of PD-L1 serves as a binding-site for SP1 and is associated with PD-L1 overexpression and increased occurrence of gastric cancer. *Cancer Immunol Immunother*. 2017;66(3):309–318. doi:10.1007/s00262-016-1936-0
52. Krawczyk P, Grenda A, Wojas-Krawczyk K, et al. PD-L1 gene copy number and promoter polymorphisms regulate PD-L1 expression in tumor cells of non-small cell lung cancer patients. *Cancer Genetics*. 2019;237:10–18. doi:10.1016/j.cancergen.2019.06.001
53. Yu Z, Li Z, Jolicoeur N, et al. Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers. *Nucleic Acids Res*. 2007;35(13):4535–4541. doi:10.1093/nar/gkm480

BJMBR
BRAZILIAN JOURNAL OF MEDICAL AND BIOLOGICAL RESEARCH
e-ISSN 1414-431X
bjournal.org

Clinical outcome of Brazilian patients with non-small cell lung cancer in early stage harboring rare epidermal growth factor receptor mutations

Journal:	*Brazilian Journal of Medical and Biological Research*
Manuscript ID	12409.R1
Manuscript Type:	Full Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Machado-Rugolo, Juliana; University of São Paulo (USP), Department of Pathology; São Paulo State University (Unesp) Baldavira, Camila; Universidade de São Paulo Faculdade de Medicina, Departamento de Patologia Gutierrez Prieto, Tabatha; Faculdade de Medicina da Universidade de São Paulo, Department of Pathology Olivieri, Eloisa Helena Ribeiro; AC Camargo Cancer Center Fabro, Alexandre; Ribeirão Preto Medical School, University of São Paulo, Pathology; PATHOLOGY - FMRP/USP Rainho, Claudia A. ; Institute of Biosciences of Botucatu - UNESP, Genetics Castelli, Erick Cruz; São Paulo State University, Department of Pathology; São Paulo State University (Unesp) Ribolla, Paulo Eduardo Martins; São Paulo State University (Unesp); State University (Unesp), Department of Biostatistics Ab'Saber, Alexandre Muxfeldt; University of São Paulo (USP), Department of Pathology Takagaki, Teresa; Universidade de Sao Paulo Faculdade de Medicina, IncorHCFMUSP-pneumologia; Universidade de Sao Paulo Faculdade de Medicina Hospital das Clinicas Instituto do Cancer, oncologia clinica Nagai, Maria A. ; University of Sao Paulo Capelozzi, Vera L. ; University of São Paulo (USP), Department of Pathology
Keywords:	epidermal growth factor receptor, uncommon mutation, next-generation sequencing, erlotinib, lung cancer
Special Sections:	Clinical Investigation/Oncology

Reviewer 1:

Comments to the Author

"This is a very interesting and necessary work since molecular data in somatic mutations in south american cancer patients is very scarce. Before publication a number of issues requires clarification".

Response. First of all, we would like to thank the Reviewer for the critical comments that allowed us to rewrite the work to make it more fluent and clearer to justify the rationale of the study and minimize major weaknesses in regard to data presentation and interpretation. We revised the manuscript based on his comments and highlighted the answers to queries in yellow, and we are re-submitting the paper along with a point-by-point response to the reviewer' comments. Two other prerogatives need to be highlighted: 1) the respect and consideration for the patients included in the study and the benefits they will bring to other patients, and 2) the urgent establishment of the genomic profile and targeted therapies may lead to improvements in therapeutic strategies and the clinical outcomes of Brazilian patients with lung cancer. The intellectual exercise carried out between authors and reviewers undoubtedly increased the scientific value of the work. Thank you very much.

Query 1. **"The 3-stage design is not clear, where\what is stage 1 for this publication. Please elaborate".**

Response. This is a very pertinent question. The reason to include patients in stage IB-IIIA was the observation that in our cohort some patients experienced a prolonged recurrence-free survival (RFS) for non-small cell lung carcinoma (NSCLC) by chest surgery alone. However, other patients with completely resected NSCLC required adjuvant chemotherapy to avoid recurrence after complete surgical resection. In addition, clinical trials demonstrated that target therapy improved RFS in epidermal growth factor receptor (EGFR) mutation-positive stage IB–IIIA disease following complete resection. In the new version of the manuscript, we modified the Introduction section to define the choice of patients with NSCLC in early stage (see below). Thank you for the comment.

Pg. 3

Introduction

However, some patients experienced a cure for lung cancer by chest surgery alone. In contrast, other patients with completely resected NSCLC may have a minimal residual disease (MRD) not identified radiographically and require adjuvant chemotherapy.

Thus, adjuvant chemotherapy is indicated to avoid recurrence in patients who underwent complete surgical resection NSCLC. For patients with stage IIA–IIIA NSCLC postoperative, the standard treatment adjuvant is cisplatin-based chemotherapy. Pooled analysis of several clinical trials has shown that adjuvant cisplatin-doublet chemotherapy improved the 5-year disease-free survival (DFS) rate by 5.8% and the 5-year OS rate by 5.4% (4). Moreover, the ADAURA trial demonstrated that osimertinib improved DFS in epidermal growth factor receptor (EGFR) mutation-positive stage IB–IIIA disease following complete resection (5).

Query 2. **"The selection of the in silico prediction tools is intriguing. Many/most of those are meant for germline genetic variants and are very good to predict loss of function, but since EGFR is an oncogene, we would like to predict gain of function also. Please consider using something like the "Cancer genome interpreter" availble at https://www.cancergenomeinterpreter.org/home, or else, and reflect on advantages/limitations of the algorithm used in the paper**".

Response. We thank the very pertinent comment, thus contributing to improve the quality of our article. We also checked the "Cancer Genome Interpreter" as recommended by the reviewer and was also entered into the M&M of the new version (please, see below). Furthermore, we reflect on the limitations of the algorithms used in the article by inserting in the discussion section of the new manuscript (please, see below). Thank you for the careful analysis of our manuscript.

Pg. 6

M&M

Variants of unknown significance (VUS) were checked on the Cancer genome interpreter (https://www.cancergenomeinterpreter.org/home) and VarSome database.

Pg. 15

Discussion

The main limitation of our research was that the comprehensive interpretation of sequence variants requires not only well-established databases but also appropriate functional analyses, such as evaluating the effect of gene expression of these found variants.

Query 3. **"Novel variants could also arise from sequencing artifacts, plsease discuss allelic frequency distribution of novel *versus* know variants and kind of transitions observed, especially the ones associated to formalin related DNA damage as described in Wong, S.Q., Li, J., Tan, A.YC. et al. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively**

parallel sequencing. BMC Med Genomics 7, 23 (2014). https://doi.org/10.1186/1755-8794-7-23)".

Response. We appreciate the comment and suggestion that assisted in the revision of the work and that certainly qualify and confer improvements in our text. To avoid this kind of artifacts, our investigation used only fresh-frozen (FF) tumors to eliminate disproportionate levels of C>T/G>A changes as seen in formalin-fixed paraffin-embedded tumors (FFPE). Furthermore, to minimize sequencing artifacts, PCR amplification or germline variants, we used as inclusion criteria for genetics variants:

(1) number of reads with the altered base in the tumor ≥10
(2) mutations detected at a position of total read depth of ≥ 100
(3) frequency of the reads with the altered base in the tumor ≥ 5% except for variants that are also reported in the COSMIC database
(4) minor allele frequency <1% in two publicly available databases (1000 Genomes, and Exome Aggregation Consortium).

This observation made by Reviewer was included in new version of the manuscript in Material and Methods section (see below). We appreciated this question very much. Thank you.

Pg. 4

Sample Collection

Since that novel variant could also arise from sequencing artifacts, especially the ones associated to formalin related DNA damage as described in Wong, S.Q et al (9), our investigation was done in fresh-frozen specimens from Brazilian patients with lung cancer, collected during surgical resection.

Pgs. 5 and 6

Identification of Single Nucleotide Variants (SNVs) and Indels

In order to reducing the effects of PCR amplification and sequencing artifacts, sequencing data analyses was performed on the Molecular Genetics and Bioinformatics Laboratory of the Experimental Research Unit (UNIPEX) at the Medical School of São Paulo State University (UNESP).

To reduce false-positive somatic mutations which might originate from germline variants were followed the following cut off criteria: number of reads with the altered base in the tumor ≥10, mutations detected at a position of total read depth of ≥ 100, frequency of the reads with the altered base in the tumor ≥ 5% except for variants that are also reported in the COSMIC database, minor allele frequency < 0.1% in two publicly available databases, namely 1000 Genomes[1], and Exome Aggregation Consortium[2].

Query 4. **"The association studies are very interesting, but because of the small sample size they are probably underpowered for the null hypothesis across the analysis. Maybe a post hoc power analysis would help the readers to asses the analysis results".**

Response. The reviewer is right again. Because of the small sample size, statistical analysis does not make sense. Therefore, we re-write the manuscript discussing the clinical outcome case by case with references to the literature, and not based on powered for significance (please, see Results in the new version of the manuscript). Thanks a lot for the comment.

Query 5. **"in Figure 1 the lolliplot is missing, instead there is a oncoprinter (in cbioportal jargon) and several bar plot depicting slightly similar data, please consider redesigning this figure**".

Response. If the Reviewer agrees, we take the liberty to maintain the Figure 1 because its purpose is only illustrative for the reader. Thank you.

Query 6. **"in table 2 there are wild type EGFR patients treated with erlotinib, is that right? 7. supp 2 is missing column headers in the pdf version of the ms and supp 1 is scrambled"**.

Response. We appreciate your careful analysis of our manuscript. As stated in Query 4, we removed all tables with statistical analysis and reported the results case-by-case. We also corrected the column headers in suppl Table 2, as well as suppl Table 1. Thank for the observation.

Reviewer: 2

Comments to the Author

Dear authors,

"The manuscript investigates a relevant and interesting topic, but the text as is

requires major revision of both English, structure and content. It is unclear trough out the manuscript the main goal of the study, the results are extensive and over descriptive of details that distract from your main points. Furthermore, much of your results are based on a fairly heterogeneous sample (age, treatment, sex, ... and other confounders) that are not accounted during your statistical analysis".

Response. First of all, we would like to thank the Reviewer for the critical comments that allowed us to rewrite the work to make it more fluent and clearer to justify the rationale of the study and minimize major weaknesses in regard to data presentation and interpretation. We revised the manuscript based on his comments and highlighted the answers to queries in yellow, and we are re-submitting the paper along with a point-by-point response to the reviewer' comments. Two other prerogatives need to be highlighted: 1) the respect and consideration for the patients included in the study and the benefits they will bring to other patients, and 2) the urgent establishment of the genomic profile and targeted therapies may lead to improvements in therapeutic strategies and the clinical outcomes of Brazilian patients with lung cancer. The intellectual exercise carried out between authors and reviewers undoubtedly increased the scientific value of the work. Thank you very much.

Section Editor: 1

Comments to the Author:

To complement the comments of reviewer 2 we are inserting the comments sent by the reviewer to the editor:

"Although the topic approached by the authors is interesting and relevant for field, the results presented are inconclusive given the high heterogeneity of their sample and the lack of adequate control in their studies. It is unclear if the authors investigated tumor or germinative mutations, which severely impact the interpretation of their results. Futhermore, the manuscript requires major rewritting. Their abstract is confusing and doesn't make clear their goal or study object. The results are overdescriptive with little information added".

Response. We would like to thank the Reviewer for the critical comments that allowed us **to rewrite the work** to make it more fluent and clearer to justify the rationale of the study and minimize major weaknesses in regard to data presentation and interpretation. We revised the manuscript based on his comments and highlighted the answers to queries in yellow, and we are re-submitting the paper along with a point-by-point response to the reviewer' comments. Two other prerogatives need to be highlighted: 1) the respect and consideration for the patients included in the study and the benefits they will bring to other patients, and 2) the urgent establishment of the genomic profile and targeted therapies may lead to improvements in therapeutic strategies and the clinical outcomes of Brazilian patients with lung cancer. The intellectual exercise carried out between authors and reviewers undoubtedly increased the scientific value of the work.

Thank you very much.

Clinical outcome of Brazilian patients with non-small cell lung cancer in early stage harboring rare epidermal growth factor receptor mutations

Juliana Machado-Rugolo[1,2], Camila Machado Baldavira[1], Tabatha Gutierrez Prieto[1], Eloisa Helena Ribeiro Olivieri[3], Alexandre Todorovic Fabro[1,4,] Claudia Aparecida Rainho[5,6], Erick Cruz Castelli[7,8], Paulo Eduardo Martins Ribolla[8,9], Alexandre Muxfeldt Ab'Saber[1], Teresa Takagaki[10], Maria Aparecida Nagai[11,12], Vera Luiza Capelozzi[1]

[1] Laboratory of Histomorphometry and Lung Genomics, Department of Pathology, Medical School, University of São Paulo (USP), São Paulo, Brazil

[2] Health Technology Assessment Center, Clinical Hospital of Botucatu, Medical School, São Paulo State University (Unesp), Botucatu, Brazil

[3]International Center of Research/CIPE, AC Camargo Cancer Center, São Paulo, São Paulo, Brazil

[4]Department of Pathology and Legal Medicine, Respiratory Medicine Laboratory, Ribeirão Preto Medical School, University of São Paulo (USP), São Paulo, Brazil

[5] Institute of Biosciences, Department of Chemical and Biological Sciences, São Paulo State University (Unesp), Botucatu, Brazil

[6]Molecular Genetics and Bioinformatics Laboratory, Experimental Research Unit (UNIPEX), Medical School, São Paulo State University (Unesp), Botucatu, Brazil

[7]Department of Pathology, Medical School, São Paulo State University (Unesp), Botucatu, Brazil

[8]Biotechnology Institute (IBTEC), São Paulo State University (Unesp), Botucatu, Brazil

[9]Institute of Biosciences, Department of Biostatistics, Plant Biology, Parasitology and Zoology, State University (Unesp), Botucatu, Brazil

[10]Division of Pneumology, Instituto do Coração (Incor), University of São Paulo Medical School (USP), São Paulo, Brazil

[11] Department of Radiology and Oncology, Medical School, University of São Paulo (USP), São Paulo, Brazil

[12] Laboratory of Molecular Genetics, Center for Translational Research in Oncology, Cancer Institute of São Paulo (ICESP), São Paulo, Brazil

Abstract

Introduction: The common epidermal growth factor receptor (EGFR) mutations, such as the L858R point mutation in exon 21 and the in-frame deletional mutation in exon 19, have been definitively associated with response to EGFR-tyrosine kinase inhibitors (EGFR-TKI). However, the clinical outcome and response to treatment for many other rarer mutations are still unclear. In this study, we report the results of Brazilian patients in stage IB–IIIA non-small cell lung cancer (NSCLC) following complete resection with minimal residual disease and *EGFR* mutations treated with adjuvant chemotherapy and/or EGFR-TKIs.

Methods: The frequency of EGFR mutations was investigated in 70 cases of NSCLC early stage. Mutations in exons 18 and 20, uncommon mutations in exons 19 and 21, as well as in exons 3, 7, 14, 16, 22, 27, and 28, and/or the presence of different mutations in a single tumor (complex mutations) was considered rare.

Results: *EGFR* mutations were detected in 23 tumors (32.9%). Fourteen cases carried rare mutations and were treated with platin-based chemotherapy and two with erlotinib. The clinical outcome was described case by case with references to the literature. Notably, we found two rare EGFR mutations and one of them with an unknown response to chemotherapy and/or EGFR-TKIs.

Conclusions: We have provided complementary information concerning the clinical outcome and treatment in patients with NSCLC early stage for several rare *EGFR* mutations not previously or only rarely reported. Description of cases harboring rare mutations can provide the decision-making procedure in this subset of patients.

Keywords: Lung cancer, next generation sequency, rare *EGFR* mutation, platin-based chemotherapy, erlotinib.

Introduction

In Brazil, the latest 2021 tumor registration data show that there were approximately 17.760 new cases of lung cancer in men and 12.440 in women (1). Non-small cell lung cancer (NSCLC) accounts for 85% of lung cancer in Brazilian patients [1]. Almost 20% of them have early-stage disease (stage I–II) and around 30% of patients have locally advanced disease (stage III) at the time of diagnosis of NSCLC (2) . Operative resection is the gold standard curative treatment for patients with NSCLC; nevertheless, local or systemic relapse of the disease is frequent despite complete resection. Five-year overall survival (OS) rates are reported to be 70% for stages IB-IIA, 50%–60% for stages IIA-IIB, and 35% for stage IIIA (3). However, some patients experienced a cure for lung cancer by chest surgery alone. In contrast, other patients with completely resected NSCLC may have a minimal residual disease (MRD) not identified radiographically and require adjuvant chemotherapy.

Thus, adjuvant chemotherapy is indicated to avoid recurrence in patients who underwent complete surgical resection NSCLC. For patients with stage IIA–IIIA NSCLC postoperative, the standard treatment adjuvant is cisplatin-based chemotherapy. Pooled analysis of several clinical trials has shown that adjuvant cisplatin-doublet chemotherapy improved the 5-year disease-free survival (DFS) rate by 5.8% and the 5-year OS rate by 5.4% (4). Moreover, the ADAURA trial demonstrated that osimertinib improved DFS in epidermal growth factor receptor (EGFR) mutation-positive stage IB–IIIA disease following complete resection (5). Therefore, the establishment of the genomic profile and targeted therapies may lead to improvements in therapeutic strategies and the clinical outcomes of lung cancer patients.

The “classical” *EGFR* mutations, such as the L858R point mutation in exon 21 and the in-frame deletional mutation in exon 19, have been undoubtedly associated with target therapy response, while many other EGFR mutations are rarely detected in patients with NSCLC, and information about their association with response to target therapy are still uncertain (6). In cell culture, rare exon 20 mutations make transformed cells less responsive to target therapy, suggesting that other mechanisms probably contribute to primary resistance in metastatic NSCLC (7). For many of the rare mutations, the effect on responsiveness remains unknown. Thus, it is of extreme importance for the clinical decision-making process to share information of patients harboring such mutations,

particularly when the outcome of epidermal growth factor receptor-tyrosine kinase receptor (EGFR-TKI) treatment is available.

Previous studies performed by our group using frozen tissue specimens and next-generation sequencing (NGS) from a cohort of 70 patients have shown that Brazilian patients represent a genetically admixed NSCLC cohort with important predictive and prognostic implications (8, 9). In the same case series and array used (QIAamp DNA Mini Kit Qiagen), we detected some Brazilian patients with NSCLC harboring rare *EGFR* variants with clinical significance to be defined.

Therefore, the aim of the present study was to explore the clinical outcome of patients with resected NSCLC in the early stage harboring rare somatic *EGFR* mutations. We also considered mutations in exons 18 and 20, uncommon mutations in exons 19 and 21, and/or the presence of different mutations in a single tumor (complex mutations) were all considered rare. We found 14 cases with rare mutations. These patients were treated with adjuvant chemotherapy and two of them with erlotinib. In this scenario, we discuss their clinical outcome case by case with references to the literature if obtainable, or *in silico* analysis. We also highlight two rare pathogenic mutations that were identified and one case showing EGFR-TKI efficacy with such a mutation. The other rare mutations have been previously described in case reports.

Material and Methods

Sample Collection

Since that novel variant could also arise from sequencing artifacts, especially the ones associated to formalin related DNA damage as described in Wong, S.Q et al (9), our investigation was done in fresh-frozen specimens from Brazilian patients with lung cancer, collected during surgical resection from August 2003 to August 2010 at the A.C. Camargo Cancer Center, a tertiary referral center for the treatment of lung cancer in São Paulo, Brazil. The histologic classification was reviewed according to the World Health Organization 2021 classification system (11) and stratified as non-squamous non-small lung cancer (46 cases) and squamous non-small cell lung cancer (n=24). *EGFR* status was associated with baseline characteristics (age, sex, ethnicity, smoking history, stage at diagnosis, histology, type of *EGFR* mutation), treatments (radiotherapy, platinum-based chemotherapy, tyrosine kinase inhibitors) and outcome (relapse-free survival [RFS] after primary surgical resection, relapse, and development of distant metastases.

The study was approved in accordance with the ethical standards of the local committee on human experimentation (Research Ethics Committee of University of São Paulo Medical School - CAAE: 79769017.1.0000.5440; opinion number: 2.673.320). Informed consent was waived because of the retrospective study design. The identity of the subjects under this retrospective analysis was also omitted and anonymized.

EGFR Mutation Analysis

Next-Generation Sequencing Analysis

The DNA of fresh tumor tissue was extracted using the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's recommendations, and quantified using the Qubit® 3.0 Fluorometer (Invitrogen, Life Technologies, Carlsbad, CA, USA). The full genomic sequence of EGFR (2 kbp upstream 5'UTR, 2 kbp downstream 3'UTR, coding regions, and intronic sequences) was targeted and captured using TruSeq Custom Amplicon Panel v1.5kit (TSCAP, Illumina, SanDiego, CA) followed by massively parallel sequencing of enriched fragments on an Illumina MiSeq platform (Illumina, San Diego, CA, USA) consisting of 150 bp paired-end reads (300 cycles). All tumor specimens had an average sequencing depth of the target region ≥100× and coverage of the target region >90% at 30×.

Identification of Single Nucleotide Variants (SNVs) and Indels

In order to reducing the effects of PCR amplification and sequencing artifacts, sequencing data analyses was performed on the Molecular Genetics and Bioinformatics Laboratory of the Experimental Research Unit (UNIPEX) at the Medical School of São Paulo State University (UNESP). Sequencing data analyses were performed on the Molecular Genetics and Bioinformatics Laboratory of the Experimental Research Unit (UNIPEX) at the Medical School of São Paulo State University (UNESP). The raw sequencing data were base-called and demultiplexed using MiSeq Reporter v.1.8.1 (Illumina, San Diego, CA, USA) with default parameters, and FastQC files were generated for downstream data analysis. Filtered reads were aligned to the human genome (hg19, GRCh37) using the Burrows-Wheeler Alignment tool (BWA) v.0.7.10. After alignment, the SAMtools software was applied to convert the alignment files to an indexed binary alignment map format. The single nucleotide variants (SNVs) and short insertions and deletions (INDELs) were called using the GATK UnifiedGenotyper,

including HaplotypeCaller with default parameters based on hg19 and annotated with dbSNP version 144[1]. To reduce false-positive somatic mutations which might originate from germline variants were followed the following cut off criteria: number of reads with the altered base in the tumor ≥10, mutations detected at a position of total read depth of ≥ 100, frequency of the reads with the altered base in the tumor ≥ 5% except for variants that are also reported in the COSMIC database, minor allele frequency < 0.1% in two publicly available databases, namely 1000 Genomes[2], and Exome Aggregation Consortium[3].

https://www.ncbi.nlm.nih.gov/SNP/

http://www.1000genomes.org/

http://exac.broadinstitute.org/

***In silico* Prediction Tools and Tumor Samples to Guidance Uncommon Profiles**

The variants were annotated using the VEP software based on the consequences, predicted impacts, and reported allele frequencies in the population. Variants of unknown significance (VUS) were checked on the Cancer genome interpreter (https://www.cancergenomeinterpreter.org/home) and VarSome database, a single website that allows access to publications, ClinVar, and all *in silico* prediction tools including the Genomic Evolutionary Rate Profiling (GERP). We also used the Mastermind bioinformatic platform to examine further clinical diagnostic variant interpretation. Next, the mutation plots were generated using the online customized MutationMapper tool at cBioPortal for Cancer Genomics. In this study, all variants were classified according to the American College of Medical Genetics and Genomics (ACMG) recommendations, which allows classification as either "pathogenic," "likely pathogenic," "likely benign," "benign" or "uncertain significance".

Data evaluation

All analyses of the cohort were exploratory. No statistical test was applied given the small size of the cohort, implying a reduced power of significance.

Results

Characteristics of the study cohort

Table 1 summarizes the baseline characteristics of the cohort. 70 Brazilian patients with NSCLC surgically resected were included in the study. The median age of the patients was 78 years (range, 41–96), with similar distribution between older and younger patients. Forty-four patients were male (62.9%), 48 (68.6%) with European ancestrally (68.6%), and 21 (30.0%) smokers. Adenocarcinoma histotype was diagnosed in 46 (65.71%) patients and squamous cell carcinoma in 24 (34.3%) patients. Classification of patients according to TNM staging included 14 individuals (20.0%) in stage IA, 10 individuals (14.3%) in stage IB, 10 individuals (14.3%) in stage IIA, 13 individuals (18.6%) in stage IIB, and 11 individuals (15.7%) in stage IIIA. During the postoperative period, pulmonary and mediastinal relapse occurred in 11 patients (15.7%) and distant metastasis (brain and bone) in 12 patients (17.4%). Of the 70 patients, 23 (32.8%) received platinum-based chemotherapy, and two patients (2.9%) received erlotinib. Thirty-eight patients (64%) died. The median follow-up was 49 months (range, 0-175).

3.2. Frequency of *EGFR* somatic mutations

In our casuistry, EGFR mutations (target exons 1-28) were detected in 23 of 70 patients (32.9%), 44 males and 21 females. Classic or sensitizing activating mutations (exons 18-21) occurred in 14 patients (20.0%), being 9 of 14 (12.9%) in exon 19 (in-frame deletion). Uncommon mutations in exons 3, 7, 14, 16, 22, 27, and 28, and/or complex mutations were identified in 9 tumors corresponding to 12.9% of the entire population analyzed. In **Tables 2** and **Table 3**, the results of NGS sequences are listed in detail. **Figure 1** depicts an illustration obtained from cbioportal jargon and bar plots with data of mutation frequency, somatic mutation variants, and mutation code.

3.3 Clinical outcome in patients harboring mutations cluster in exons 18-21

Exon 18 Mutations

G719C (exon 18) + S768I (exon 20)

Patient report. The G719 + S7681 pathogenic double mutation was found in a 70-year-old male smoker with European ancestrally and adenocarcinoma showing micropapillary characteristics in stage IIIA. After local recurrence, the patient was treated with a combination chemotherapy of CDDP and etoposide. The patient died 13 months after the surgical resection.

Reference. The complex mutation has been previously described in a few reports (12,13). *In vitro* data suggest that the coexistence of the E709A + G719C mutation confers resistance to adjuvant treatment, including EGFR-TKIs (14). Another report did not confirm the low sensitivity to adjuvant treatment of the double mutant (15).

Comment. According to the cases described so far, the *G719C + S768I* complex mutation seems to be associated with resistance to adjuvant treatment including EGFR-TKI.

Exon 19 Mutations

p.E746_R748del/p.A750P

Patient report. The p.E746_R748del/p.A750P pathogenic complex mutation was found in two patients female nonsmokers patients, 90-year-old and 96-year-old, both with European ancestrally, and adenocarcinoma stage IIIA with acinar and solid characteristics. After pulmonary and mediastinal local relapse, both patients were treated with carboplatin and Taxol, which was interrupted early due to toxicity. A clinically meaningful improvement of symptoms was obtained within a few days and a partial response of tumor assessment after 6 weeks. After 9 months, both patients presented disease progression and died.

Reference. The presence of p.E746_R748del/p.A750P has already been reported in one Caucasian patient by Jakobsen and colleagues (16) but without specific information on the efficacy of EGFR-TKI.

Comment. Similarly, from the single case reported in the literature showing progression following gefitinib, the two patients treated in our institution had a rapid symptomatic improvement but died 9 months after disease progression.

Exon 20 Mutations

L813R + R776C

Patient report. The L813R + R776C EGFR mutation was found in a 70-year-old current heavy male smoker, with European ancestrally and adenocarcinoma stage IIIA with

acinar characteristics. After pulmonary local relapse, the patient was treated with CDDP/Navelbine. A tumor progression was observed after 10 weeks at the first treatment assessment. Erlotinib was subsequently given at the starting dose of 150 mg daily. The patient experienced 30 months of progression-free survival.

Reference. The *L813R + R776C* mutation was previously described by Ham and colleagues (17) in a patient with squamous NSCLC, who benefited from afatinib and experienced 11 months of progression-free survival. The same mutation has been found by Moran and colleagues (18), but in association with the L861 in exon 20, in a never-smoking patient with adenocarcinoma with lepidic features. In this case, the patient had a complete and long-lasting remission after afatinib treatment.

Comment. Similarly, from the case reported by Moran where the L861 was linked to *R776C*, in our experience the *L813R + R776C* EGFR mutation was associated with response to erlotinib.

Exon 21 Mutations

P848L

Patient report. The P848L mutation was found in **a** 94-year-old current heavy male smoker, with European ancestrally and squamous cell carcinoma with non-keratinizing features. After lobectomy, the tumor pathological stage was classified as IB. During the first three months of follow-up, the patient developed pulmonary relapse and bone metastases. He was treated with radiotherapy and a combination of CDDP/Navelbine, followed by erlotinib with rapid tumor progression, and died.

Reference. P848L has been described in 16 carcinoma patients in the Catalogue of Somatic Mutations in Cancer (COSMIC) (19), representing a frequency of about 0.6% in cancer patients (20). A recent study showed that the P848L germline variant of EGFR contributes to cellular transformation and predicts resistance to TKIs, and patients harboring P848L likely benefitting from JAK inhibitors (21). Furthermore, very rare exon 21 mutations have shown to be more sensitive to afatinib (second-generation EGFR-TKIs) and osimertinib (third-generation EGFR-TKIs) than first-generation EGFR-TKIs in preclinical studies (22).

Comment. This is an unusual case of a patient with a tumor in stage IB harboring a P848L mutation with no response to adjuvant treatment, including TKI, and rapid progression of the disease.

3.4 Clinical outcome of patients harboring rare *EGFR* mutations with uncertain significance

Exon 3 Mutations

M137V

Patient report. M137V mutation was found concurrently with other two unknown mutations, c.1881-462C>T (intron15) and A1165V (exon 27), in a 78-year-old female non-smoker with European ancestrally and adenocarcinoma showing lepidic characteristics. The patient underwent pneumectomy and the tumor was staged as IB. She received adjuvant radiotherapy and a combination of CDDP with Navelbine with partial response of the tumor and is alive after 8 years of follow-up.

Reference. The M137V mutation and the c.1881-462C>T and A1165V variants are reported in the dbSNP database (rs754854319) without any data regarding their clinical significance. Chakroborty et al. (2019) described this alteration in colon adenocarcinomas from human samples available in the GENIE database (23). VarSome database predicted their significance to be uncertain. The DEOGEN2, FATHMM-MKL, LIST-S2, M-CAP, MutationAssessor, MutationTaster, and SIFT computational analysis inferred pathogenicity, while BayesDel_addAF, DANN, EIGEN, MVP, and PrimateAI tools) predicted benignity. The GERP score (5.13) indicates that the Met amino acid in position 137 lies in a strongly conserved region, further deposing in favor of predicting pathogenicity.

Comment. Our patient with adenocarcinoma stage IB harboring M137V mutation responded partially to adjuvant treatment and is alive after 8 years of follow-up.

Exon 7

A289V Mutation

Patient report. The A289V mutation was found in a 94-year-old male patient, with European ancestrally and adenocarcinoma with solid characteristics, and staged as IIB. The patient died two months after the surgical resection.

Reference. The A289V mutation in exon 7 encodes an amino acid of the extracellular domain of the EGFR protein and had been found only in the glioblastoma, low-grade glioma, head, and neck neoplasms, being rarely described in NSCLCs (24, 25) . Dai et al. (2018) were the first to present a case report of NSCLC in a Chinese patient with *EGFR* p.A289V mutation, whose corresponding treatment showed only a partial response to

first-generation EGFR-TKIs (25). The A289V mutation was classified as likely pathogenic by ACMG. VarSome database analysis showed a pathogenic computational result based on ten pathogenic predictions from BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MutationAssessor, MutationTaster, and SIFT *vs.* two benign predictions from MVP and PrimateAI.

Comment. The M137V mutation seems to occur in patients with advanced age in the early stage of the disease rapidly progressing to death.

Exon 8.

C326F Mutation

Patient report. C326F mutation was detected in a stage IIB keratinizing squamous cell carcinoma from a smoker patient 89-year-old, who underwent pneumonectomy and died one month after.

Reference. This genomic alteration had been described in a case report of Lhermitte-Duclos disease, also known as dysplastic cerebellar gangliocytoma, which has been shown to be an activation mutation resulting in increased *EGFR* autophosphorylation and indicating high receptor activation (26). VarSome's *in silico* analysis of the C326 mutation revealed 11 pathogenic predictions from BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MVP, MutationAssessor, MutationTaster, and SIFT *vs.* one benign prediction from PrimateAI.

Comment. The C326F also occurred in a patient with advanced age with squamous cell carcinoma in an early stage disease rapidly progressing to death.

Exon 14.

V547I-G569S Mutations.

Patient report. The V5471-G569S mutation, concomitantly with a point mutation R776C, was found in a 69-year-old male patient with African ancestrally and non-keratinizing squamous cell carcinoma, stage IIB, who presented pulmonary relapse 11 months after surgical resection. The patient was treated with radiotherapy and a combination of carboplatin and Taxol but died due to brain metastases.

Reference. VarSome database analysis revealed seven pathogenic predictions for the V5471-G569S mutation according to DANN, EIGEN, EIGEN PC, FATHMM-MKL, FATHMM-XF, LRT, MutationAssessor, and MutationTaster against 12 benign predictions from BayesDel addAF, BayesDel noAF, DEOGEN2, FATHMM, LIST-S2,

MVP, MetaLR, MetaSVM, PrimateAI, REVEL, and SHIFT 4G. The G569S mutation with concomitant point mutation R776C in the VarSome computational analysis led to a benign computational result based on 12 benign predictions from BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MVP, MutationAssessor, MutationTaster, PrimateAI and SIFT *vs.* no pathogenic predictions.

Comment. In our case, the V5471-G569S mutations were detected in non-keratinizing squamous cell carcinoma, stage IIB, from a patient of African ancestry who did not respond to adjuvant chemotherapy and died after brain metastases.

Exon 16.

R635W Mutation

Patient report. The mutation R635W was found in a 64-year-old female non-smoker patient with adenocarcinoma, stage IIA, who received adjuvant chemotherapy and radiotherapy and died three years after the surgery.

Reference. The mutation R635W is reported in the dbSNP database (rs369399038) of uncertain significance. *In silico* analysis predicted a deleterious effect based on 11 pathogenic predictions from BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MutationAssessor, MutationTaster, PrimateAI, and SIFT vs. one benign prediction from MVP.

Comment. Another patient from our Institution with adenocarcinoma stage IIA harboring the mutation R635W experienced partial response to radiotherapy and chemotherapy with rapid progression of the disease.

Exon 22

D871N Mutation

Patient report. The D871N mutation was found in a 69-year-old heavy smoker male patient with African ancestrally and keratinizing squamous cell carcinoma, stage IIB, who underwent pneumectomy. The tumor relapsed in the central nervous system. The patient received a combination of carboplatin and Taxol and died eleven months after.

Reference. The VarSome database produced a final likely pathogenic computational result based on seven pathogenic predictions from DANN, DEOGEN2, FATHMM-MKL, LIST-S2, M-CAP, MutationTaster, and SIFT *vs.* five benign predictions from BayesDel_addAF, EIGEN, MVP, MutationAssessor, and PrimateAI.

Comment. Another uncommon mutation harbored in squamous cell carcinoma from patient with poor tumor response.

Exon 28 and Exon 29

Non-coding mutations

Patient report. Three male patients older than 70 years with IIA adenocarcinoma harboring respectively c.*582T>G (rs869064669), c.*708T>C (rs751311059), c.*934G>A (rs988454507), and c.*1060G>A (rs771422383) who underwent lobectomy and received a combination of CDDP with Navelbine are alive nine years after surgical resection.

Reference. The four VarSome analyses resulted in benign computational results, with one benign prediction from DANN *vs.* no pathogenic prediction in all four cases. Moreover, all four variants had a GERP score of less than 5.5, an indication that the position of these variants was not highly conserved (-0.878, 3.56, -2.5, and -0.396, respectively).

Comment. It is worth noting, however, that the use of more sensitive sequencing techniques might reveal rare mutants at a higher frequency.

Discussion

Previous studies from the present case series have shown that Brazilian patients represent a genetically admixed NSCLC cohort with clinical implications. In fact, we found that patients with adenocarcinoma harboring somatic mutations of *EGFR*, *CTLA4*, *PDCD1LG2*, or *ZEB2* that only underwent surgical treatment and developed brain metastases may have the worst prognosis (8,9). Regarding *EGFR*, we detected approximately 13% of NSCLC Brazilian patients harboring *EGFR* somatic mutations with clinical significance to be defined. These findings motivated us to re-analyze *EGFR* mutations in the 70-case series of NSCLC. Because of the small sample size, we discuss their clinical outcome case by case with references to the literature, and not based on powered for significance.

Under the objectives of the present study, we found: (1) rare mutations, uncommon mutations, and/or complex somatic mutations identified in 20% of Brazilian population analyzed with prevalence of European ancestrally, non-smoker male older than 77 yrs, adenocarcinoma histology stage IB-IIB; (2) rare *G719C (exon 18) + S768I*

(exon 20) complex somatic mutation seems to be associated with resistance to combination chemotherapy of CDDP and etoposide with short PFS; (3) rare *p.E746_R748del/p.A750P* complex *somatic* mutations resistant to carboplatin and Taxol with short PFS; (4) rare *L813R + R776C (exon 20*) complex somatic mutation associated with response to erlotinib and prolonged PFS; (5) rare *P848L (exon 21*) somatic mutation resistant to combination of CDDP/Navelbine and erlotinib with short PFS; (6) uncommon *M137V (exon 3) and non-coding (exons 28 and 29*) somatic mutations with response to combination of CDDP with Navelbine and prolonged PFS; (7) uncommon somatic mutations *A289V (exon 7), C326F (exon 8), V547I-G569S (exon 14), R635W (exon 16), D871N (exon 22)* with resistance to combination of carboplatin and Taxol and short PFS.

The frequency of *EGFR* mutations in NSCLC ranging from 5 to 20%, depending on the cohort studied (6). Erlotinib and gefitinib are small molecules that target tumor *EGFR* mutation and are effective to treat patients with this subgroup of tumors, showing a response rate of approximately 70% (27). Two classical activating mutations represent the most common mutations: a short in-frame deletion of exon 19 and a point mutation (CTG to CGG) in exon 21 at nucleotide 2573 leading to the replacement of leucine by arginine at codon 858 (L858R). However, other *EGFR* mutations do not effectively cause *EGFR* tyrosine-kinase activity and may be associated with acquired TKIs resistance, therefore affecting the sensitivity of the EGFR-TKIs (28). In the case of rare EGFR mutations, the literature is very limited in terms of clinical outcome and response to treatment and may underpower the value of target therapy.

Therefore, from an *EGFR* mutation screen, performed on our patients, we detected NSCLC with rare *EGFR* mutations and illustrated the response rate after chemotherapy and erlotinib treatment. In detail, we analyzed 70 tumors and identified 14 cases, with rare mutations, defined as mutations in exons 18, 20, and 21 uncommon mutations in exons 3, 7, 14, 16, 22, 27, 28, and/or complex mutations (different mutations present in a single tumor). These mutations accounted for approximately 20% of all the cases screened coinciding with other studies in South America including Brazil (29) and Colombia (30) but was lower than those found in Asia and higher than those of European populations (31, 32). The current cohort included patients from all regions of the country who came to a tertiary referral center for the treatment of lung cancer. These patients present different education degrees, socioeconomic status and ancestrally characterizing a genetic admixture, inherited from European, African, and Asian immigrants (31, 32). As expected, we observed a high prevalence of rare and uncommon *EGFR* mutation in

Brazilian patients with European ancestry, compared to those found in the native European population (33). Twenty-three of our patients were treated with platin-based chemotherapy and two patients received erlotinib, and we report herein the individual case reports regarding the literature and *in silico* analysis. We observed that patients with rare *EGFR* mutations could receive platinum-based chemotherapy as a first-line treatment, due to their low response rates and short PFS in response to erlotinib. The clinical outcome observed supported the different predictive values of the single *EGFR* mutations in terms of treatment effectiveness. Of note, in our cohort, exon 20 mutations were associated with erlotinib sensitivity, and the patient experienced 30 months of progression-free survival; however, there are only a few reports in the literature of cases responding to EGFR-TKIs. Therefore, we infer that the evaluation of the single mutations, case by case, could be useful also in the presence of exon 20 mutations, as in the presence of other uncommon *EGFR* mutations.

Although we showed the potential of NGS technique in small cohort of NSCLC, future validation is necessary using a similar cohort with a large set of patients to corroborate the results observed in this tumor tissue type. The main limitation for our exploratory analysis is the small number of NSCLC cases used but it was minimized by the data obtained using this technique. The present study is largely descriptive and exploratory, and extension of our findings are essential. Another limitation of our research was that the comprehensive interpretation of sequence variants requires not only well-established databases but also appropriate functional analyses, such as evaluating the effect of gene expression of these found variants.

In summary, we have provided complementary information concerning the effectiveness of chemotherapy and erlotinib in patients with NSCLC early stage for several rare *EGFR* mutations not previously or only rarely reported. This data together with similar studies in the literature and *in silico* analysis reinforce the decision-making process in such subsets of patients.

References

1. de Sá VK, Coelho JC, Capelozzi VL, de Azevedo SJ. Lung cancer in Brazil: epidemiology and treatment challenges. *Lung Cancer (*Auckl*)* 2016 Nov 14;7:141-148. doi: 10.2147/LCTT.S93604.

2. Araujo LH, Baldotto C, Castro G Jr, Katz A, Ferreira CG, Mathias C, et al. Lung cancer in Brazil. *J Bras Pneumol* 2018 Jan-Feb;44(1):55-64. doi: 10.1590/S1806-37562017000000135.

3. Monteiro AS, Araújo SRC, Araujo LH, Souza MC. Impact of microvascular invasion on 5-year overall survival of resected non-small cell lung cancer. *J Bras Pneumol* 2022 Jul 8;48(3):e20210283. doi: 10.36416/1806-3756/e20210283.

4. Booth CM, Shepherd FA, Peng Y, Darling G, Li G, Kong W, et al. Adjuvant chemotherapy for non-small cell lung cancer: practice patterns and outcomes in the general population of Ontario, Canada. *J Thorac Oncol* 2012 Mar;7(3):559-66. doi: 10.1097/JTO.0b013e31823f43af.

5. Zhang SS, Ou SI. Deconstructing ADAURA: It is Time to Forgo Adjuvant Platinum-Based Chemotherapy in Resected IB-IIIA EGFR+ NSCLC (Except with RB Alterations?) When Adopting Adjuvant Osimertinib. *Lung Cancer (*Auckl*)* 2022 Apr 26;13:23-31. doi: 10.2147/LCTT.S358902.

6. Riely GJ, Pao W, Pham D, Li AR, Rizvi N, Venkatraman ES, et al. Clinical course of patients with non-small cell lung cancer and epidermal growth factor receptor exon 19 and exon 21 mutations treated with gefitinib or erlotinib. *Clin Cancer Res* 2006 Feb 1;12(3 Pt 1):839-44. doi: 10.1158/1078-0432.CCR-05-1846.

7. Greulich H, Chen TH, Feng W, Jänne PA, Alvarez JV, Zappaterra M, et al. Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. *PLoS Med* 2005 Nov;2(11): e313. doi: 10.1371/journal.pmed.0020313.

8. Parra ER, Jiang M, Machado-Rugolo J, Yaegashi LB, Prieto T, Farhat C, et al. Variants in Epithelial-Mesenchymal Transition and Immune Checkpoint Genes Are Associated With Immune Cell Profiles and Predict Survival in Non-Small Cell Lung Cancer. *Arch Pathol Lab Med* 2020 Oct 1;144(10):1234-1244. doi: 10.5858/arpa.2019-0419-OA.

9. Machado-Rugolo J, Gutierrez Prieto T, Fabro AT, Parra Cuentas ER, Sá VK, Baldavira CM, et al. Relevance of PD-L1 Non-Coding Polymorphisms on the Prognosis of a Genetically Admixed NSCLC Cohort. *Pharmgenomics Pers Med* 2021 Feb 15; 14:239-252. doi: 10.2147/PGPM.S286717.

10. Wong SQ, Li J, Tan AY, Vedururu R, Pang JM, Do H, Ellul J, Doig K, Bell A, MacArthur GA, Fox SB, Thomas DM, Fellowes A, Parisot JP, Dobrovic A; CANCER 2015 Cohort. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. BMC Med Genomics. 2014 May 13;7:23. doi: 10.1186/1755-8794-7-23.

11. Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, et al. The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. J Thorac Oncol 2022;17(3):362-387. doi: 10.1016/j.jtho.2021.11.003.

12. Chantharasamee J, Poungvarin N, Danchaivijitr P, Techawatanawanna S. Clinical outcome of treatment of metastatic non-small cell lung cancer in patients harboring uncommon EGFR mutation. *BMC Cancer* 2019 Jul 17;19(1):701. doi: 10.1186/s12885-019-5913-9.

13. Bi H, Ren D, Wu J, Ding X, Guo C, Miura S, et al. Lung squamous cell carcinoma with rare epidermal growth factor receptor mutation G719X: a case report and literature review. *Ann Transl Med* 2021 Dec;9(24):1805. doi: 10.21037/atm-21-6653.

14. Tam IY, Leung EL, Tin VP, Chua DT, Sihoe AD, Cheng LC, et al. Double EGFR mutants containing rare EGFR mutant types show reduced in vitro response to gefitinib compared with common activating missense mutations. *Mol Cancer Ther* 2009 Aug;8(8):2142-51. doi: 10.1158/1535-7163.MCT-08-1219

15. Hijiya N, Miyawaki M, Kawahara K, Akamine S, Tsuji K, Kadota J, et al.. Fosphorylation status of epidermal growth factor receptor is closely associated with responsiveness to gefitinib in pulmonary adenocarcinoma. *Hum Pathol* 2008 Mar;39(3):316-23. doi: 10.1016/j.humpath.2007.10.027.

16. Jan Nyrop Jakobsen, Eric Santoni-Rugiu, Morten Grauslund, Linea Melchior, Jens Benn Sørensen. Concomitant driver mutations in advanced EGFR-mutated non-small-cell lung cancer and their impact on erlotinib treatment. *Oncotarget* 2018 May 25; 9(40): 26195–26208.

17. Han C, Ding X, Li M, Luo N, Qi Y, Wang C. Afatinib, an effective treatment for patient with lung squamous cell carcinoma harboring uncommon EGFR G719A and R776C co-mutations. *J Cancer Res Clin Oncol* 2022 May;148(5):1265-1268. doi:10.1007/s00432-021-03864-4.

18. Moran T, Taus A, Arriola E, Aguado C, Dómine M, Rueda AG, et al.; Study Group for the Uncommon EGFR Mutations in Spain. Clinical Activity of Afatinib in Patients With Non-Small-Cell Lung Cancer Harboring Uncommon EGFR Mutations: A Spanish Retrospective Multicenter Study. *Clin Lung Cancer* 2020 Sep;21(5):428-436.e2. doi: 10.1016/j.cllc.2020.04.011.

19. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 2019; 47: D941-D947. doi: 10.1093/nar/gky1015.

20. Santamaría I, Menéndez ST, Balbín M. EGFR L858R mutation may go undetected because of P848L in cis mutation. *J Clin Oncol* 2013; 31: e420-1. doi: 10.1200/JCO.2012.47.3512.

21. Sarcar B, Gimbrone NT, Wright G, Remsing Rix LL, Gordian ER, Rix U, et al. Characterization of epidermal growth factor receptor (EGFR) P848L, an unusual EGFR variant present in lung cancer patients, in a murine Ba/F3 model. *FEBS Open Bio* 2019; 9:1689-1704. doi: 10.1002/2211-5463.12702.

22. Zhang H. Three generations of epidermal growth factor receptor tyrosine kinase inhibitors developed to revolutionize the therapy of lung cancer. *Drug Des Devel Ther* 2016; 10:3867-3872. doi: 10.2147/DDDT.S119162.

23. Chakroborty D, Kurppa KJ, Paatero I, Ojala VK, Koivu M, Tamirat MZ, et al. An unbiased in vitro screen for activating epidermal growth factor receptor mutations. *J Biol Chem* 2019; 294:9377-9389. doi: 10.1074/jbc.RA118.006336.

24. An Z, Aksoy O, Zheng T, Fan QW, Weiss WA. Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies. *Oncogene* 2018; 37:1561-1575. doi: 10.1038/s41388-017-0045-7.

25. Dai L, Su X, Lu L, Lv D. Nonsmall cell lung cancer with rare exon 7 p.A289V mutation in the EGFR gene responds to Icotinib treatment: A case report. *Medicine* (Baltimore) 2018; 97: e13809. doi: 10.1097/MD.0000000000013809.

26. Colby S, Yehia L, Niazi F, Chen J, Ni Y, Mester JL, et al. Exome sequencing reveals germline gain-of-function EGFR mutation in an adult with Lhermitte-Duclos disease. *Cold Spring Harb Mol Case Study* 2016; 2: a001230. doi: 10.1101/mcs.a001230.

27. Jackman DM, Holmes AJ, Lindeman N, Wen PY, Kesari S, Borras AM, et al. Response and resistance in a non-small-cell lung cancer patient with an epidermal growth factor receptor mutation and leptomeningeal metastases treated with high-dose gefitinib. *J Clin Oncol* 2006 Sep 20;24(27):4517-20. doi: 10.1200/JCO.2006.06.6126.

28. Balak MN, Gong Y, Riely GJ, Somwar R, Li AR, Zakowski MF, et al. Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. *Clin Cancer Res* 2006 Nov 1;12(21):6494-501. doi: 10.1158/1078-0432.CCR-06-1570.

29. Freitas HC, Torrezan GT, da Cunha IW, Macedo MP, Karen de Sá V, Corassa M, et al. Mutational Portrait of Lung Adenocarcinoma in Brazilian Patients: Past, Present, and Future of Molecular Profiling in the Clinic. *Front Oncol* 2020;10:1068. doi: 10.3389/fonc.2020.01068.

30. Arrieta O, Cardona AF, Corrales L, Campos-Parra AD, Sánchez-Reyes R, Amieva-Rivera E, et al.; CLICaP. The impact of common and rare EGFR mutations in response to EGFR tyrosine kinase inhibitors and platinum-based chemotherapy in patients with

non-small cell lung cancer. *Lung Cancer* 2015 Feb;87(2):169-75. doi: 10.1016/j.lungcan.2014.12.009.

31. Rahman S, Kondo N, Yoneda K, Takuwa T, Hashimoto M, Orui H, et al. Frequency of epidermal growth factor receptor mutations in Bangladeshi patients with adenocarcinoma of the lung. *Int J Clin Oncol* 2014; 19(1):45-9. doi: 10.1007/s10147-012-0515-4.

32. Smits AJ, Kummer JA, Hinrichs JW, Herder GJ, Scheidel-Jacobse KC, Jiwa NM, et al. EGFR and KRAS mutations in lung carcinomas in the Dutch population: Increased EGFR mutation frequency in malignant pleural effusion of lung adenocarcinoma. *Cell Oncol* (Dodr) 2012; 35: 189-196. doi: 10.1007/s13402-012-0078-4.

33. De Pas T Toffalorio F, Manzotti M, Fumagalli C, Spitaleri G, Catania C, et al. Activity of epidermal growth factor receptor-tyrosine kinase inhibitors in patients with non-small cell lung cancer harboring rare epidermal growth factor receptor mutations. *J Thorac Oncol* 2011 Nov;6(11):1895-901. doi: 10.1097/JTO.0b013e318227e8c614.

Funding

This work was supported by São Paulo Research Foundation (FAPESP; 2018/20403-6, 2019/12151-0), the National Council for Scientific and Technological Development (CNPq; 483005/2012-6), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES; Finance Code 001).

Footnotes

1. https://www.ncbi.nlm.nih.gov/SNP/
2. http://www.1000genomes.org/
3. http://exac.broadinstitute.org/

Abbreviations

3' UTR, 3'untranslated region; ACMG, American College of Medical Genetics and Genomics; BWA, Burrows-Wheeler Alignment, CHT, chemotherapy; CI, confidence interval; COSMIC, Catalogue of Somatic Mutations in Cancer; DNA, deoxyribonucleic

acid; EGFR-TKIs, EGFR tyrosine kinase inhibitors; *EGFR*, Epidermal Growth Factor Receptor; GERP, Genomic Evolutionary Rate Profiling; HR, hazard ratio; IASLC, International Association for the Study of Lung Cancer; INDELS, insertions and deletions; NGS, Next Generation Sequencing; NSCLC, non-small cell lung carcinomas; OS, overall survival; SNV, Single Nucleotide Variants; VUS, variants of uncertain significance.

Acknowledgments

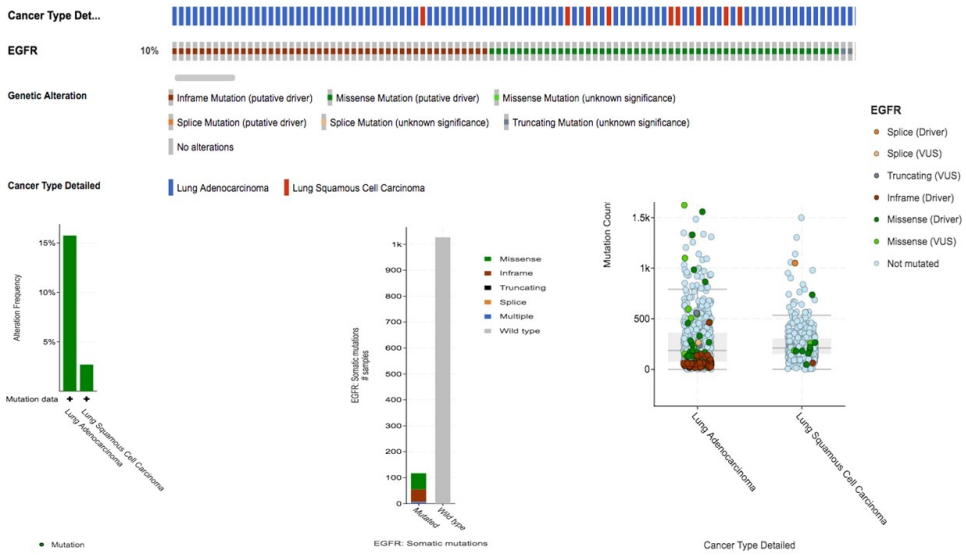
We appreciate all subjects who participated in this study and the Illumina members for assistance with the initial runs.

Author Contributions

All authors contributed equally to this study. Conception and design: VLC and JM-R; Writing, review, and editing: VLC, JM-R, CMB and TGP; Data analysis and interpretation: VLC, JM-R, CMB, TGP; Statistical analysis: VLC and TGP; Provision of study materials or patients: CAR, ECC, PEMR, AMAS, TYT, MAN, ERPC; Administrative support: VLC; Final approval of manuscript: All authors.

Legends of figures

Figure 1. Lollipop plots showing the distribution of EGFR variants in exons 18-21(**A**) and other EGFR uncommon variants (**B**) in a Brazilian NSCLC cohort as identified by Next-Generation Sequencing. The plots were obtained by the informatic tool Mutation Mapper—cBioPortal for Cancer Genomics.



Lollipop plots showing the distribution of EGFR variants in exons 18-21(A) and other EGFR uncommon variants (B) in a Brazilian NSCLC cohort as identified by Next-Generation Sequencing. The plots were obtained by the informatic tool Mutation Mapper—cBioPortal for Cancer Genomics.

338x190mm (96 x 96 DPI)

Table 1. Frequency of demographic and clinical characteristics of 70 NSCLC patients.

Characteristics	Number of Patients (N=70)
Age, years	
Median range, (41-96)	77 (41-96)
≤77	32 (45.7%)
>77	33 (47.1%)
Gender	
Male	44 (62.9%)
Female	21 (30.0%)
Ancestrally	
European	48 (68.6%)
Asian	2 (2.9%)
African	2 (2.9%)
Smoking status	
Smoker	21 (30.0%)
Non-smoker	7 (10.0%)
Histological subtype	
Adenocarcinoma	46 (65.71%)
Squamous cell carcinoma	24 (34.3%)
Pathological TNM stage†	
IA	14 (20.0%)
IB	10 (14.3%)
IIA	10 (14.3%)
IIB	13 (18.6%)
IIIA	11 (15.7%)
Relapse	
No	13 (18.6%)
Locoregional	11 (15.7%)
Distant metastasis	12 (17.4%)
Post operative treatment	
Chemotherapy platinum-based	23 (32.8%)
Radiotherapy	4 (4.3%)
Chemoradiotherapy	6 (8.6%)
Tyrosine kinase inhibitor (erlotinib)	2 (2.9%)
Status for overall survival	
Live	23 (32.9%)
Dead	38 (54.3%)
Follow-up (months)	49 (0-175)
***EGFR* status**	
Classical mutation (Exon 18-21)	14 (20.0%)
Uncommon mutation (Exon 3, 7, 14, 16, 22, 27, 28)	9 (12.9%)
Wild type	47 (67.1%)

Abbreviations: NSCLC, non-small cell lung cancer. Some cases lacked follow-up information: gender (5); age (5); race (18); smoking status (42); TNM stage (6); Relapse (36); Chemotherapy (47); Radiotherapy (43), Status (9). † 8th Edition International Association for the Study of Lung Cancer.

Table 2: Spectrum of *EGFR* variants in exons 18-21 identified in a Brazilian NSCLC cohort by Next-Generation Sequencing

Exon/Intron	ID rs	HGVS Nucleotide	HGVS Protein	Variant Type	Frequency	ACMG Classification
18	rs28929495	c.2155G>T	G719C (p.Gly719Cys)	Missense	1	Pathogenic
19	rs121913229	c.2248G>C	A750P (p.Ala750Pro)	Missense	7	Likely Pathogenic
19	rs121913421	c.2235_2249del	E746_A750del (p.Glu746_Ala750del)	inframe_deletion	5	Pathogenic
19	rs121913436	c.2239_2247delTTAAGAGAA	L747_E749del (p.Leu747_Glu749del)	inframe_deletion	1	Likely Pathogenic
19	rs121913442	c.2240_2254del	L747_T751del (p.Leu747_Thr751del)	inframe_deletion	1	Likely Pathogenic
20	rs121913465	c.2303G>T	S768I (p.Ser768Ile)	Missense	1	Pathogenic
20	rs121434568	c.2438T>G	L813R (p.Leu813Arg)	Missense	2	Pathogenic
20	rs1275022697	c.2326C>T	R776C (p.Arg776Cys)	Missense	1	Likely Pathogenic
21	rs148934350	c.2543C>T	P848L (p.Pro848Leu)	Missense	1	Pathogenic

Abbreviations: EGFR, Epidermal Growth Factor Receptor; NSCLC, *Non-small Cell Lung Cancer*; rs, Reference Single Nucleotide Polymorphism; HGVS, Human Genome Variant Society; ACMG, American College of Medical Genetics and Genomics.

Table 3. Spectrum of uncommon *EGFR* variants identified in a Brazilian NSCLC cohort by Next-Generation Sequencing.

Exon/Intron	ID rs	HGVS Nucleotide	HGVS Protein	Variant Type	Frequency	ACMG Classification
3	rs754854319	c.409A>G	M137V (p.Met137Val)	Missense	3	Uncertain Significance
7	rs149840192	c.866C>T	A289V (p.Ala289Val)	Missense	1	Likely Pathogenic
8	rs886037891	c.977G>T	C326F (p.Cys326Phe)	Missense	1	Likely Pathogenic
14	rs144943614	c.1639G>A	V547I (p.Val547Ile)	Missense	1	Uncertain Significance
14	rs779076899	c.1705G>A	G569S (p.Gly569Ser)	Missense	1	Uncertain Significance
15	rs538888597	c.1881-462C>T	*	Intron	1	Uncertain Significance
16	rs369399038	c.1903C>T	R635W (p.Arg635Trp)	Missense	1	Uncertain Significance
22	rs376822837	c.2611G>A	D871N (p.Asp871Asn)	Missense	1	Uncertain Significance
27	rs35918369	c.3494C>T	A1165V (p.Ala1165Val)	Missense	1	Uncertain Significance
28	rs869064669	c.*582T>G	*	3_prime_UTR_variant	1	Uncertain Significance
28	rs751311059	c.*708T>C	*	3_prime_UTR_variant	1	Uncertain Significance
28	rs988454507	c.*934G>A	*	3_prime_UTR_variant	1	Uncertain Significance
28	rs771422383	c.*1060G>A	*	3_prime_UTR_variant	1	Uncertain Significance

Abbreviations: EGFR, Epidermal Growth Factor Receptor; NSCLC, *Non-small Cell Lung Cancer*; rs, Reference Single Nucleotide Polymorphism; HGVS, Human Genome Variant Society; ACMG, American College of Medical Genetics and Genomics.