

# Análise Estatística da Expressão do Gene CD274 em Tecidos com Câncer

## Statistical Analysis of CD274 Gene Expression in Tissues with Cancer

Odilon J. dos Santos<sup>1</sup>

<sup>1</sup>Instituto Metrópole Digital  
Universidade Federal do Rio Grande do Norte (UFRN)  
Natal – RN – Brasil  
Disciplina: PROJETOS EM BIOINFORMÁTICA  
Orientadora: Profa. Dra. Tirzah Braz Petta Lajus

odilonjulio@ufrn.edu.br

**Abstract.** *This is the analysis of the possible correlation between gene expression of the PD-L1 (CD274) gene in cancer tissues, more specifically, in this case, cancer of the skin and lungs of about 500 individuals and their respective after diagnosis.*

**Resumo.** *Trata-se da análise da possível correlação existente entre a expressão gênica do gene PD-L1 (CD274) e seus fatores de transcrição em tecidos com câncer, mais especificamente, neste caso, câncer de pele e de pulmão, de cerca de 500 indivíduos e a suas respectivas sobrevidas, após diagnosticados.*

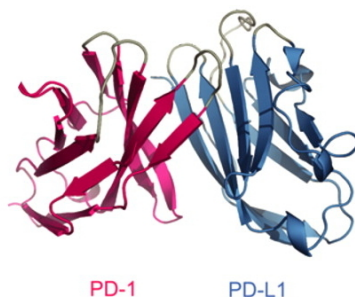


Figure 1. Proteína PD-L1

**1 Introdução** O ligante de morte programada 1 (PD-L1), também denominado como B7-H1, é uma proteína que, em humanos, é codificada pelo gene CD274. Estudos especulam que ele exerce papel fundamental na supressão do sistema imunológico, causando a redução da proliferação das células T, que são os glóbulos brancos responsáveis pela defesa do organismo diante de antígenos.

Dentre as possibilidades de tratamento de alguns tipos de câncer, segue emergente a terapia baseada na tentativa de regular a expressão gênica do PD-L1. Desta forma,

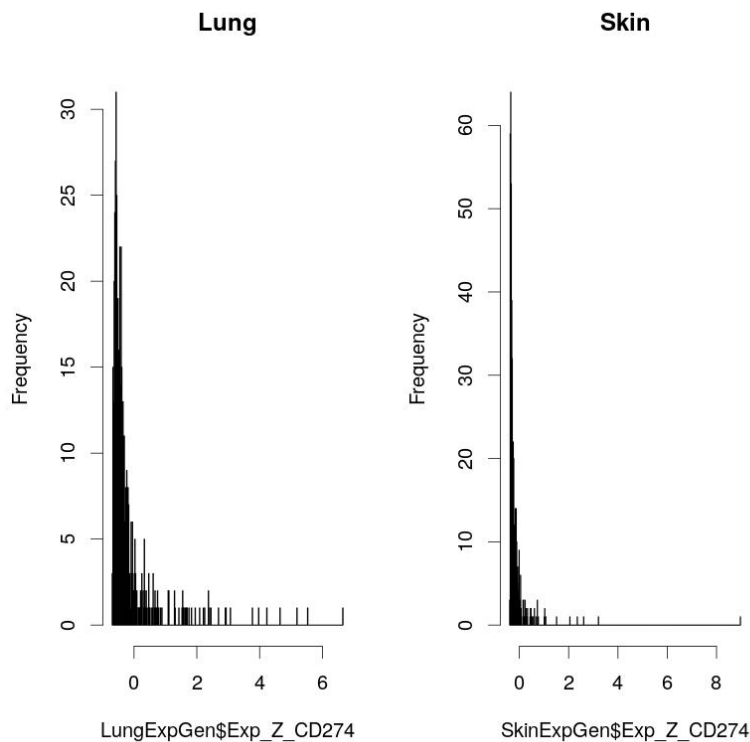
**2 Objetivos** Diante da justificativa acima, o objetivo central dessa análise inicial é quantificar o quão importante é, para o tratamento de alguns tipos de câncer, o controle ou regulação da expressão gênica do gene CD274 por meio de inibidores (anti-PDL1).

[illegible]

Especificamente, e paralelamente, foram analisados os dois tecidos com câncer, acima referidos. Mas pesquisas futuras estão direcionadas aos demais tecidos e, além disso, às demais doenças ou eventos específicos do organismo.

Com intuito de permitir melhor avaliação e estudo dos dados, foi utilizado o Software R e suas principais bibliotecas com fim criar um *script* padronizado que, futuramente, pode ser utilizado para avaliar outros pacientes.

Continuando a análise, foram captados os indivíduos que representam os 5% com maior expressão gênica e os 5% com menor expressão. Tais indivíduos tiveram uma análise secundária, onde foi correlacionada a sobrevida de cada um, em meses, deles e seus respectivos valores (em Z-Score) de expressão do gene em tela. Após tal refino, foi possível conjecturar que há, de fato, em números, uma relação direta entre o gene CD274 estar superexpresso e a baixa sobrevida dos pacientes.



**Figure 3. Histogramas da Expressão Gênica do CD274**

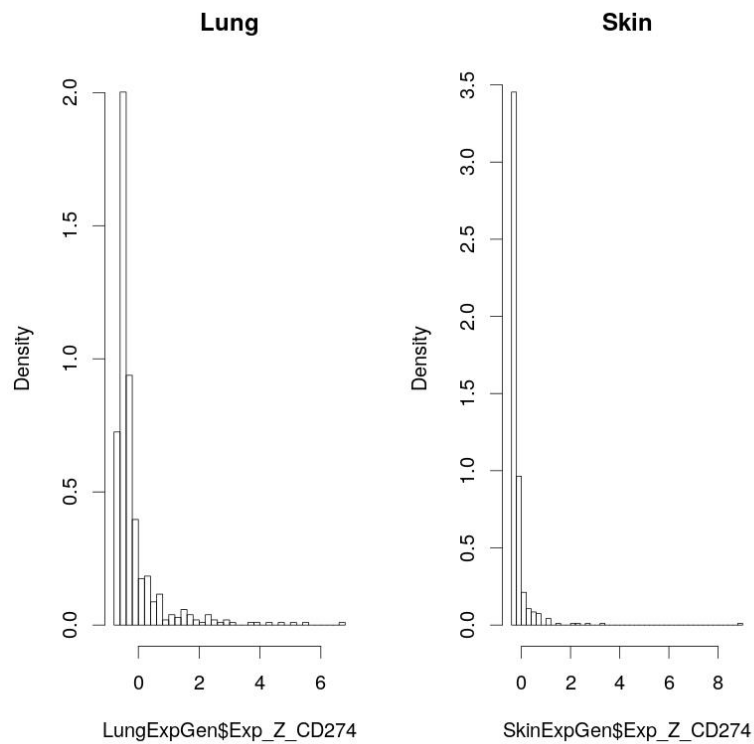
No entanto, se fez necessária a busca por dados relativos aos fatores de transcrição do CD274, bem como pela relação estatística entre eles e o próprio gene referido. A partir daí, uma análise numérica para encontrar os fatores de transcrição (TFs) que exercem uma forte correlação (positiva ou negativa) com o CD274 demonstraram que, dentre os cerca de 800 TFs, apenas pouco mais de três dezenas deles pertenciam ao grupo de interesse, juntamente dados de ambos os tecidos (pulmão e pele).

Com isso, fazendo referência ao mesmo grupo de pacientes e mesmos tecidos, foi analisado um novo dado: a expressão gênica de cada fator de transcrição. Diante de um grande número de caminhos possíveis de serem seguidos, decidimos avaliar somente a expressão dos TFs nos pacientes que apresentaram maior e menor expressão gênica do CD274, a fim de verificar mais detalhadamente tal correlação.

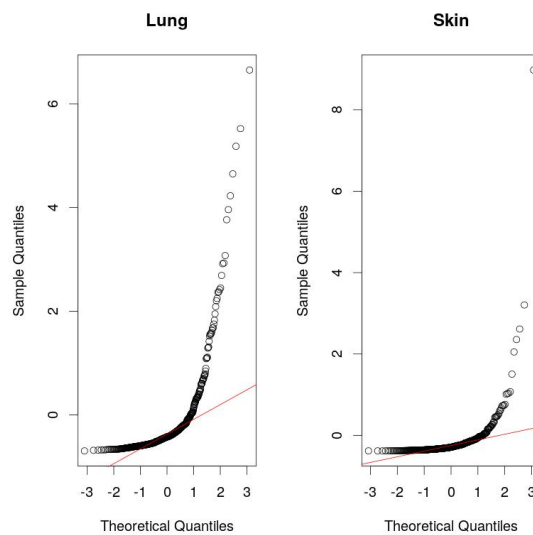
Desta forma, fazendo verificações similares às que foram inicialmente feitas com o gene CD274, chegamos à conclusão estatística, diante de um p-valor abaixo de 5%, que há uma forte correlação entre a sobrevida dos pacientes e superexpressão de um único gene: BCL3. Mesmo alterando filtros e refazendo os testes, apenas o fator de transcrição citado demonstrou essa correlação.

Diante da expectativa de serem encontrados mais genes (TFs) que possuam sua expressão gênica fortemente relacionada à sobrevida de pacientes, resta buscar futuramente novos dados e repetir os testes com intuito de diminuir o leque de apresentadas neste estudo preliminar.

É importante lembrar que esse estudo foi realizado com uma banco de dados muito



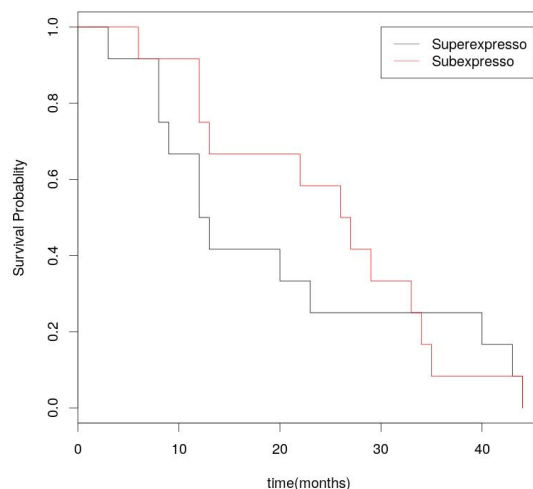
**Figure 4. Histogramas da Expressão Gênica do CD274 normalizado.**



**Figure 5. Teste que constata que a distribuição da expressão gênica não segue uma distribuição normal.**

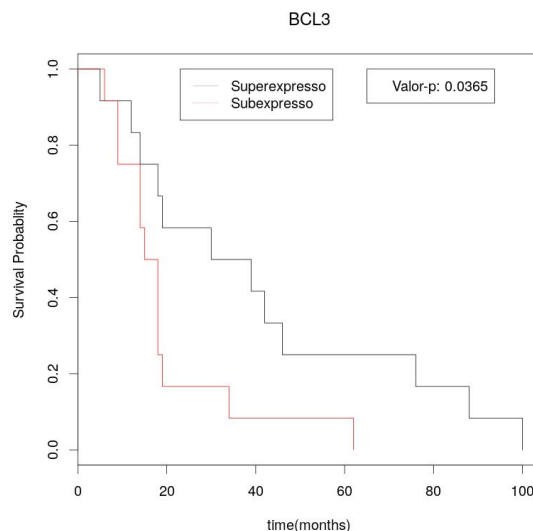
pequeno e se faz necessário um maior aprofundamento nas análises. Contudo, ainda será realizado um teste de hipótese estatística para reafirmar ou confrontar a informação encontrada até então.

Desta forma, essa investigação, pelo pequeno número de amostras até então avali-

[illegible][illegible]

adas, ainda é inconclusiva, mas ascende a expectativa real de que a relação aqui encontrada tem grande probabilidade de também existir numa população com maior número de indivíduos.

Por fim, vale salientar que foram avaliados pacientes que ainda possuíam o câncer e/ou faleceram em decorrência da doença, já que apenas ficou perceptível que o gene pode



**Figure 9. Curva de Kaplan-Meier da Análise de Sobrevida do Fator de Transcrição BCL3.**

estar superexpresso em indivíduos livres da doença.

Há, então, uma proposta de criação de script padronizado ou software que faça tais análises estatísticas, tratando o dado e verificando a expressão de genes sua relação com a sobrevida do paciente com câncer, apresentando, por fim, a curva de subbrevida do Kaplan-Meier.

Tal análise foi possível, diante do auxílio principal do Prof. Tetsu Nakamoto (Instituto Metr pole Digital - UFRN) e da bolsista de inicia  o cient fica, Jheyne La na, do BioMe (Bioinformatics Multidisciplinary Environment - UFRN), os quais ajudaram, tanto fornecendo dados e informa  es, como sugerindo m todos e testes estat sticos para uma verifica  o mais detalhada.

Mais informa  es que justificam e embasam essa an lise podem ser encontradas em (Brahmer et al. 2012), (Reck et al. 2016), (Iwai et al. 2002) e outros que facilmente podem ser encontrados em conte do p blico.

## References

- [Brahmer et al. 2012] Brahmer, J. R., Tykodi, S. S., Chow, L. Q., Hwu, W.-J., Topalian, S. L., Hwu, P., Drake, C. G., Camacho, L. H., Kauh, J., Odunsi, K., et al. (2012). Safety and activity of anti-pd-11 antibody in patients with advanced cancer. *New England Journal of Medicine*, 366(26):2455–2465.
- [Iwai et al. 2002] Iwai, Y., Ishida, M., Tanaka, Y., Okazaki, T., Honjo, T., and Minato, N. (2002). Involvement of pd-11 on tumor cells in the escape from host immune system and tumor immunotherapy by pd-11 blockade. *Proceedings of the National Academy of Sciences*, 99(19):12293–12297.
- [Reck et al. 2016] Reck, M., Rodr guez-Abreu, D., Robinson, A. G., Hui, R., Cs szi, T., F l p, A., Gottfried, M., Peled, N., Tafreshi, A., Cuffe, S., et al. (2016). Pembrolizumab versus chemotherapy for pd-11–positive non–small-cell lung cancer. *New England Journal of Medicine*, 375(19):1823–1833.