

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
PROGRAMA DE PÓS-GRADUAÇÃO EM DOENÇAS INFECCIOSAS**

CARLOS HENRIQUE DETTMANN FANTECELLE DE CASTRO

**ASSINATURA TRANSCRICIONAL DA IMUNOSSENECÊNCIA EM LESÕES
DE PACIENTES COM LEISHMANIOSE CUTÂNEA CAUSADA POR
*LEISHMANIA BRAZILIENSIS***

**VITÓRIA/ES
2020**

CARLOS HENRIQUE DETTMANN FANTECELLE DE CASTRO

**ASSINATURA TRANSCRICIONAL DA IMUNOSSENESCÊNCIA EM LESÕES
DE PACIENTES COM LEISHMANIOSE CUTÂNEA CAUSADA POR
*LEISHMANIA BRAZILIENSIS***

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Doenças Infecciosas do Centro de Ciências da Saúde da Universidade Federal do Espírito Santo, como requisito final para a obtenção do grau de Mestre em Doenças Infecciosas.

Orientadores: Prof. Dr. Daniel Cláudio de Oliveira Gomes e Dra. Luciana Polaco Covre

VITÓRIA/ES

2020

Ficha catalográfica disponibilizada pelo Sistema Integrado de Bibliotecas - SIBI/UFES e elaborada pelo autor

D278a de Castro, Carlos Henrique Dettmann Fantecelle, 1993 -
Assinatura transcricional da imunossenescência em lesões de pacientes com leishmaniose cutânea causada por *Leishmania braziliensis* / Carlos Henrique Dettmann Fantecelle de Castro. - 2020.
132 f. : il.

Orientador: Daniel Cláudio de Oliveira Gomes.

Coorientadora: Luciana Polaco Covre.

Dissertação (Mestrado em Doenças Infecciosas) - Universidade Federal do Espírito Santo, Centro de Ciências da Saúde.

1. Senescência celular. 2. Expressão gênica. 3. Imunopatologia. 4. Leishmaniose. 5. *Leishmania braziliensis*. I. Gomes, Daniel Cláudio de Oliveira. II. Covre, Luciana Polaco. III. Universidade Federal do Espírito Santo. Centro de Ciências da Saúde. IV. Título.

CDU: 61



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Centro de Ciências da Saúde
Programa de Pós-Graduação em Doenças Infecciosas
PARECER ÚNICO DA COMISSÃO JULGADORA DE DISSERTAÇÃO DE
MESTRADO

O(a) mestrando(a) Carlos Henrique Dettmann Fantecelle de Castro apresentou a tese intitulada "Transcriptional signature of immunosenescence in lesions of patients with cutaneous leishmaniasis caused by Leishmania braziliensis" em sessão pública, como requisito final para obtenção do título de Mestre em Doenças Infecciosas do Programa de Pós-Graduação em Doenças Infecciosas do Centro de Ciências da Saúde da Universidade Federal do Espírito Santo.

Considerando a apresentação oral dos resultados, sua qualidade e relevância, a Comissão Examinadora decidiu (x) aprovar () reprovar a dissertação habilitando Carlos Henrique Dettmann Fantecelle de Castro a obter o Grau de Mestre(a) em Doenças Infecciosas.

Vitória, 17 de agosto de 2020.

Prof. Dr. Daniel Claudio de Oliveira Gomes
Universidade Federal do Espírito Santo – Orientador

Profa. Dra. Walderez Ornelas Dutra
Universidade Federal de Minas Gerais – Titular Externo

Prof. Dr. Fausto Edmundo Lima Pereira
Universidade Federal do Espírito Santo – Titular Interno



AGRADECIMENTOS

Agradeço, primeiramente, à minha família. Minha mãe Ivanete que sempre se esforçou para que eu tivesse todas as oportunidades possíveis. À minha irmã Lívia, a gente que se aguenta. Sem vocês eu não seria nada.

A todos os meus amigos, pela convivência, carinho e apoio durante todos esses anos, especialmente nesse momento tão delicado. Em especial: Luíza, Júlia, Lívia Maria, Emanuela, Igor, Margarida, Matheus, Pedro, Eduardo, Renzo, Arthur, Kaíque, Carla e Dâmaris pelos momentos compartilhados em presença física antes da quarentena, virtual durante a mesma e o apoio incondicional em todos os momentos. Eduardo, obrigado por todos os Cheems compartilhados.

Um agradecimento especial ao Luccas e ao Gian pelo apoio, companhia e pelos incontáveis “cultos” durante esse período de quarentena.

Agradeço a todo pessoal do laboratório, Carlinha, Renan, Lorenzo, Luciana e Stefani, pela companhia, apoio, conversa e discussões científicas. Obrigado Renan e Lorenzo por todos os almoços compartilhados.

A toda a equipe do NDI e da UFES, por terem mantido a infraestrutura necessária ao bom funcionamento do ambiente de trabalho no decorrer desses anos. A todos os professores que fizeram parte da minha formação durante o período do Mestrado, e também aos que fizeram parte da minha banca de avaliação.

A professora Sílvia e em especial ao professor Jaime, por terem me incentivado a começar em uma área que hoje é tão importante para mim.

Ao professor Daniel por ter me aceitado como aluno em um momento crítico, por ter confiado em mim para realizar esse projeto e por todo apoio e orientação durante a realização dessa pesquisa. À Luciana novamente, por sempre ter me ajudado imensamente e por também ter me orientado nesse trabalho.

A FAPES e a CAPES, pela bolsa e pelo apoio financeiro que permitiu a realização desse projeto, e ao Newton Fund e ao Medical Research Council – UK também pelo suporte financeiro.

A todos, meu mais sincero obrigado.

“Existe uma teoria que diz que, se um dia alguém descobrir exatamente para que serve o Universo e porque ele está aqui, ele desaparecerá instantaneamente e será substituído por algo ainda mais estranho e inexplicável. Existe uma segunda teoria que diz que isso já aconteceu.”

(O Restaurante no Fim do Universo, *Douglas Adams*, p. 5)

SUMÁRIO

1. Introdução	19
1.1 Aspectos gerais das leishmanioses	20
1.2 A leishmaniose cutânea localizada	20
1.3 Resposta imunológica e a Imunopatogênese da LTA	21
1.4 A senescência celular	23
1.5 Senescência, infecções e a patogênese da LCL	27
1.6 Sequenciamento de RNA e expressão gênica	29
1.7 A expressão gênica na LCL.....	31
2. Justificativa.....	33
3. Objetivos	35
4. Materiais e Métodos	37
4.1 Obtenção dos dados.....	38
4.2 Descrição resumida do estudo que deu origem aos dados analisados	38
4.3 Softwares utilizados	39
4.4 Análise dos dados de RNA-Seq	39
4.4.1 Preparação dos Dados.....	39
4.4.2 Análise de componentes principais	40
4.4.3 Análise de expressão gênica diferencial	40
4.4.4 Agrupamento hierárquico	41
4.4.5 Análise funcional	41
4.4.6 Deconvolução de células imunológicas.....	42
4.4.7 Correlação entre genes e populações celulares	42
4.4.8 Visualização dos resultados.....	42
4.5 <i>Script</i> e informações de sessão.....	43
5. Resultados	44
5.1 Expressão gênica diferencial	45
5.2 Análise de componentes principais	46
5.3 Análise funcional.....	47
5.4 Análise de genes relacionados à senescência	48

5.4.1 Moléculas Coestimulatórias e de <i>homing</i>	52
5.4.2 Receptores de citocinas	52
5.4.3 Moléculas adaptadoras de sinalização.....	55
5.4.4 Receptores de células NK	55
5.4.5 Controle de proliferação e ciclo celular	55
5.4.6 Sinalização	59
5.4.7 Fatores de transcrição.....	59
5.4.8 Transdução de sinal em linfócitos T	62
5.4.9 Citotoxicidade.....	62
5.4.10 Ligantes de células NK.....	62
5.4.11 SASP	66
5.5 Deconvolução de células imunológicas	69
6. Discussão.....	72
7. Conclusão	80
Referências	82

LISTA DE FIGURAS

- Figura 1 – Custo para o sequenciamento do genoma humano por ano.** Nota-se um grande declínio a partir do início de 2008, onde houve a introdução das tecnologias de NGS. Fonte: <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>.30
- Figura 2 – Gráfico de vulcão dos genes diferencialmente expressos.** O eixo y representa o valor de p, onde a linha tracejada horizontal equivale a um p ajustado de 0,05. O eixo x representa a alteração multiplicativa dos genes, em escala \log_2 , onde a linha tracejada vertical representa alteração equivalente à 1 (o dobro da expressão). Pontos azuis representam a regulação negativa, enquanto pontos vermelhos representam regulação positiva. Os genes identificados fazem parte dos genes de interesse analisados posteriormente nesse estudo.....46
- Figura 3 – Análise de Componentes Principais das amostras estudadas, contendo todos os genes analisados.** Indivíduos saudáveis (CS, em azul, n = 7) estão representados à esquerda, enquanto indivíduos infectados (LCL, em vermelho, n = 21) estão representados à direita.....47
- Figura 4 – Análise de sobre representação de ontologia gênica.** No gráfico, o tamanho dos pontos reflete a proporção de genes em relação aos genes diferencialmente expressos analisados na ontologia, e o eixo x representa o número de genes sobre representados na categoria disposta no eixo y.48
- Figura 5 – Representação em dendrograma do agrupamento hierárquico das amostras do estudo.** O agrupamento foi realizado através da diferença de expressão dos genes de interesse selecionados entre indivíduos controle (HS, n = 7) e pacientes com LCL (CL, n = 21). O eixo y representa a dissimilaridade entre os grupos, calculada através do método de ligação completa sobre as distâncias euclidianas entre as amostras.....49
- Figura 6 – Mapa de calor da expressão gênica, comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21).** No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os

genes regulados positivamente. Os genes foram agrupados conforme as categorias predefinidas e explicitas no Apêndice B. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$51

Figura 7 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de moléculas costimulatórias e de *homing* comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$53

Figura 8 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de receptores de citocinas comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$54

Figura 9 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de moléculas adaptadoras de sinalização (p ajustado $< 0,05$) comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$56

Figura 10 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de receptores de células NK comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$57

Figura 11 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de controle de proliferação e ciclo celular comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$58

Figura 12 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de sinalização comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$60

Figura 13 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de fatores de transcrição comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$61

Figura 14 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de transdução de sinal em linfócitos T comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$63

Figura 15 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de citotoxicidade comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano

representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$64

Figura 16 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de ligantes de células NK comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$65

Figura 17 – Representação em mapa de calor dos genes pertencentes ao SASP comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$67

Figura 18 – Representação em diagrama de caixa com pontos dos genes pertencentes ao SASP comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$68

Figura 19 – Mapa de calor das pontuações relativas das populações células. Os tipos celulares foram identificados pelo programa ImmQuant para deconvolução de tipos celulares do sistema imunológico, comparando o grupo controle (CS, n = 7) e o grupo de pacientes (LCL, n = 21).69

Figura 20 – Mapa de correlação entre genes selecionados e populações celulares relativas identificadas através do programa ImmQuant. Foi utilizado o teste de correlação de Spearman. Valores de p estão representados, onde azul indica p negativo e vermelho indica p positivo. ρ , Coeficiente de correlação de postos de Spearman ou Rô de Spearman. *p $< 0,05$; **p $< 0,01$; ***p $< 0,001$71

LISTA DE TABELAS

Tabela 1 – Descrição da análise de expressão gênica diferencial. *	
Estatisticamente significantes, valor de p ajustado < 0.05; ** % em relação ao número inicial de genes.....	45

LISTA DE APÊNDICES

Apêndice A – Categorias adicionais sobrerrepresentadas na ontologia gênica	94
Apêndice B – Resultado da Análise Diferencial dos genes de interesse por categorias	95
Apêndice C – Resultado da Análise de deconvolução celular dos indivíduos controle	102
Apêndice D – Resultado da Análise de deconvolução celular dos pacientes	104
Apêndice E – <i>Script</i> da análise e informações de sessão do R	106

LISTA DE ABREVIATURAS E SIGLAS

ATM	Ataxia-Telangiectasia Mutada, do inglês: <i>Ataxia-Telangiectasia Mutated</i>
ATR	Proteína Relacionada à ATM e RAD3, do inglês: <i>ATM and RAD3 Related</i>
cDNA	DNA complementar
CLA	Antígeno linfocitário cutâneo, do inglês: <i>Cutaneous Lymphocyte Associated Antigen</i>
CMV	<i>Cytomegalovirus</i>
CS	Controles Saudáveis
DC	Células Dendríticas
DDR	Resposta de Dano ao DNA, do inglês: <i>DNA Damage Response</i>
EBV	<i>Esptein-Barr virus</i>
FOX	Fatores de transcrição <i>Forkhead Box</i>
FOXO	Fatores de transcrição <i>Forkhead Box</i> classe O
GDE	Genes Diferencialmente Expressos
GEO	<i>Gene Expression Omnibus</i>
GO	Ontologia Gênica, do inglês: <i>Gene Ontology</i>
HIV	<i>Human Immunodeficiency Virus</i>
IFN- γ	Interferon gama
IL	Interleucina
<i>Inflammaging</i>	Estado proinflamatório do envelhecimento, do inglês: <i>Proinflammatory Status of Aging</i>
KIR	Receptor de célula NK semelhante a imunoglobulina, do inglês: <i>killer-cell immunoglobulin-like receptor</i>
KLR	Receptor de célula NK semelhante à lectina, do inglês: <i>Killer Cell Lectin-like Receptors</i>
LCD	Leishmaniose Cutânea Difusa

LCL	Leishmaniose Cutânea Localizada
LMC	Leishmaniose Mucocutânea
LTA	Leishmaniose Tegumentar Americana
LV	Leishmaniose Visceral
MAPK	Proteína Quinase Ativada por Mitógeno, do inglês: <i>mitogen activated protein kinase</i>
NCBI	<i>National Center for Biotechnology Information</i>
NGS	<i>Next-Generation Sequencing</i>
NK	Células <i>natural killer</i>
p-AMPK	Proteína quinase fosforilada ativada por AMP, do inglês: <i>phosphorylated AMP-activated protein kinase</i>
PC	Componente Principal, do inglês: <i>Principal Component</i>
PCA	Análise de Componentes Principais, do inglês: <i>Principal Component Analysis</i>
PCR	Reação em Cadeia da Polimerase, do inglês: <i>Polymerase Chain Reaction</i>
RNA-seq	Sequenciamento de RNA
ROS	Espécies Reativas de Oxigênio
SASP	Fenótipo Secretor Associado à Senescência, do inglês: <i>Senescence-Associated Secretory Phenotype</i>
TCR	Receptor de Células T, do inglês: <i>T Cell Receptor</i>
T _{EMRA}	Linfócitos T de memória efetora que reexpressam o marcador CD45Ra
T _{H1}	Células T <i>helper</i> tipo 1
T _{H2}	Células T <i>helper</i> tipo 2
TNF- α	Fator de Necrose Tumoral Alfa, do inglês: <i>Tumour Necrosis Factor Alpha</i>
VST	<i>Variance Stabilizing Transformation</i>
γ H2Ax	Proteína H2Ax fosforilada

RESUMO

Imunossenescência é um termo usado para descrever múltiplas características observadas no compartimento de células imunológicas. Recentemente, nosso grupo demonstrou que células T e NK senescentes circulantes se acumulam em pacientes com leishmaniose cutânea localizada (LCL), uma doença causada por parasitos do gênero *Leishmania*. O acúmulo destas populações com alta atividade inflamatória e citotóxica também foi observado nas lesões cutâneas de pacientes, onde foram correlacionadas com a gravidade da doença. Dada a importância desse fenômeno, esse trabalho teve como objetivo analisar o perfil de expressão de genes associados à senescência e à imunopatogênese da LCL. Para isso, reanalisamos dados de sequenciamento de RNA de domínio público, em que são comparadas a expressão gênica na pele de indivíduos saudáveis com a pele lesionada de indivíduos infectados por *L. braziliensis*. Nossos resultados demonstram a presença de um extensivo fenótipo senescente nas lesões de pacientes com LCL. Esse fenótipo é dominado por alterações em genes associados ao controle do ciclo celular (aumento em p16 e p21); genes de sinalização (aumento em ATM e Sestrin 2); receptores de células NK (aumento nos genes das famílias KIR, KLR e NKp); fatores de transcrição (aumento em *ZEB2* e *TBX21*, redução dos genes *FOXO*); além do aumento na maioria dos genes ligados ao fenótipo secretor associado a senescência (SASP) e de receptores de citocinas pró-inflamatórias. Demonstramos também o aumento nas frequências relativas estimadas das populações de células NK e T CD8⁺ altamente diferenciadas (T_{EM} e T_{EMRA}), as quais foram positivamente correlacionadas com os principais genes de citotoxicidade (granzimas, perforina), resposta pró-inflamatória (IL-12R, IL-15, IL-15R, IL-18R, IL-1A, IL-1β, IL-2R, IFN-γ e TNF-α) ou de processos relacionados à indução ou manutenção das características de senescência (ATF, ATM, p16, KLRG1, Sestrin 2, TRAILR2, TRAILR3). Neste trabalho, identificamos pela primeira vez a assinatura transcricional da imunossenescência nas lesões de pacientes com LCL, extrapolando nossos achados prévios que indicam seu papel na imunopatogênese. No geral este estudo melhora nossa compreensão da imunidade local e sistêmica em relação a LCL e fornece uma perspectiva para novas estratégias terapêuticas.

Palavras-chave: Leishmaniose cutânea localizada. *Leishmania braziliensis*. Imunossenescência. Fenótipo secretor associado a senescência.

ABSTRACT

Immunosenescence is a word used to describe multiple characteristics observed at the immune cells compartment. Recently, our group demonstrated that circulating senescent T and NK cells accumulate in patients with localized cutaneous leishmaniasis (LCL), a disease caused by parasites of the *Leishmania* genus. These populations with high inflammatory and cytotoxic activity were also found within cutaneous lesions of patients and correlated with disease severity. Given the importance of this phenomenon, this work aimed to analyze the expression profile of genes associated to senescence and the immunopathogenesis of LCL. For that, we reanalyzed RNA sequencing data of public domain, in which the gene expression of healthy skin biopsies is compared with skin lesion biopsies from patients infected with *L. braziliensis*. Our results demonstrate the presence of an extensive senescent phenotype on lesions of patients with LCL. This phenotype is dominated by alterations in genes associated to cell cycle control (increase of p16 and p21); signalling genes (increase of ATM and Sestrin 2); NK cell receptors (increase of genes from the KIR, KLR and NKp families); transcription factors (increase of *ZEB2* and *TBX21*, decrease of *FOXO* genes); in addition to the increase on the majority of genes belonging to the senescence-associated secretory phenotype (SASP) and receptors of proinflammatory cytokines. We also demonstrated an increase in estimated relative frequencies of highly differentiated NK and CD8⁺ T (T_{EM} and T_{EMRA}) cell subsets which were positively correlated with the main genes of cytotoxicity (granzymes, perforin), proinflammatory response (IL-12R, IL-15, IL-15R, IL-18R, IL-1A, IL-1β, IL-2R, IFN-γ and TNF-α) or processes related to the induction and maintenance of senescence features (ATF, ATM, p16, KLRG1, Sestrin 2, TRAILR2, TRAILR3). On this work, we identified for the first time an immunosenescence transcriptional signature on lesions of patients with LCL, extrapolating our previous findings that indicate its role on the immunopathogenesis of this disease. Overall, this study improves our understanding of local and systemic immunity relating to LCL and provides a perspective for new therapeutic strategies.

Key words: Localized cutaneous leishmaniasis. Leishmania braziliensis.
Immunosenescence. Senescence-associated secretory phenotype.

1. INTRODUÇÃO

1.1 ASPECTOS GERAIS DAS LEISHMANIOSES

As leishmanioses são consideradas doenças tropicais negligenciadas, com mais de um bilhão de pessoas em risco em todo o mundo (WHO, 2017). São transmitidas pelas fêmeas hematófagas de insetos flebotomíneos do gênero *Lutzomyia* (Ordem Diptera, Família Psychodidae, Subfamília Phlebotominae) e causadas por parasitos pertencentes ao gênero *Leishmania*, cujas principais espécies são: *Leishmania (Viannia) braziliensis*, *L. (V.) guyanensis* e *L. (L.) amazonensis*. (MINISTÉRIO DA SAÚDE, 2017; WHO, 2019). A doença pode apresentar diferentes manifestações clínicas, exibindo desde uma forma tegumentar (LTA) a uma forma visceral (LV). A LTA pode ainda se manifestar em 3 formas principais: a Leishmaniose Cutânea Localizada (LCL), Cutânea Difusa (LCD) e Mucocutânea (LMC) (SCORZA; CARVALHO; WILSON, 2017). Essas diferentes manifestações clínicas das leishmanioses não dependem apenas da espécie infectante e de suas características, sendo também influenciadas por fatores do hospedeiro, destacando-se resposta imunológica. Desta, o equilíbrio entre as respostas imunológicas proinflamatórias e antinflamatórias é determinante para o desfecho da infecção (SCORZA; CARVALHO; WILSON, 2017; SCOTT; NOVAIS, 2016).

Estima-se que ocorram entre 600.000-1.000.000 novos casos de LCL e LMC a cada ano mundialmente (WHO, 2017). O Brasil está entre os dez países endêmicos, os quais são responsáveis por 70% da incidência global dos casos de LCL (WHO, 2016), e no continente Americano, é considerado o país com maior número de casos (WHO, 2019). Em 2017, foram relatados 17.528 casos de LCL e de LMC (WHO, 2019). Em 2018, esse número foi de 16.432, dos quais 15.632 (95,1%) correspondiam à forma cutânea da doença (OPAS, 2019). Na região sudeste, houveram 1.646 casos de LTA no ano de 2018, dos quais 52 foram registrados no Espírito Santo (SINAN, 2019).

1.2 A LEISHMANIOSE CUTÂNEA LOCALIZADA

A LCL é uma das possíveis manifestações da LTA, e é considerada a forma mais comum da doença (ALVAR et al., 2012; SALMAN; RUBEIZ; KIBBI, 1999), a qual pode ser causada por diferentes espécies de *Leishmania*, dentre elas a *Leishmania*

braziliensis (SCORZA; CARVALHO; WILSON, 2017). A LCL se manifesta após um período de incubação assintomático como uma lesão papular ou nodular única (ou em poucos números), que se desenvolve no local da picada do flebotomíneo. Em alguns casos, principalmente na infecção por *L. braziliensis*, pode-se apresentar uma linfadenopatia como indicação inicial de infecção que precede o aparecimento da lesão cutânea (SCORZA; CARVALHO; WILSON, 2017). Com a evolução da doença, as lesões iniciais podem se tornar ulceradas, com fundo granulomatoso, bordas elevadas e bem delimitadas, onde são encontrados poucos parasitos. Essas características são de grande importância para a observação clínica e diagnóstico da LCL (SALMAN; RUBEIZ; KIBBI, 1999; SCORZA; CARVALHO; WILSON, 2017). A doença evolui de forma crônica, levando à cicatrização das lesões na grande maioria dos casos após o tratamento, o que também pode ocorrer de forma espontânea em casos mais raros (COSTA et al., 1990).

1.3 RESPOSTA IMUNOLÓGICA E A IMUNOPATOGÊNESE DA LTA

A resposta à infecção por parasitos do gênero *Leishmania* apresenta uma amplitude que varia desde uma forte resposta celular, caracterizada principalmente por uma hipersensibilidade tardia e alta produção de interferon gama (IFN- γ), à uma resposta com altos níveis de anticorpos, mas que não apresentam hipersensibilidade tardia (CARVALHO et al., 1994; SCOTT; NOVAIS, 2016).

Em geral, a resposta imunológica durante a LCL é majoritariamente celular e caracterizada por uma inflamação crônica mediada por células T, envolvendo uma forte resposta do tipo T *helper* tipo 1 (T_{H1}) mediada por células T CD4⁺ e T CD8⁺ as quais são produtoras de IFN- γ , fator de necrose tumoral alfa (TNF- α), granzimas e perforina (BACELLAR et al., 2002; CARVALHO et al., 2012; LOUZIR et al., 1998; RIBEIRO-DE-JESUS et al., 1998; SCOTT; NOVAIS, 2016).

Linfócitos T CD4⁺, após o reconhecimento de antígenos podem se desenvolver em diversos subtipos auxiliares (*helper*), entre eles os subtipos T_{H1} e T_{H2} (MURPHY; WEAVER, 2017). Na resposta adaptativa ao parasito, linfócitos T CD4⁺ do tipo T_{H1} são essenciais para o controle da infecção e são recrutados para as lesões da pele onde produzem IFN- γ e TNF- α , citocinas essenciais para promoção da ativação de macrófagos para eliminação do parasito (KAYE; SCOTT, 2011; SCOTT; NOVAIS,

2016). A forma ativa da doença cutânea caracteriza-se pela presença de uma resposta mista entre T_H1 e T_H2 , marcada por altos níveis das citocinas $IFN-\gamma$, $TNF-\alpha$, interleucina (IL)-12, IL-4 e IL-10. A cura clínica está associada com uma resposta T_H1 com alta produção de $IFN-\gamma$ principalmente por células T $CD4^+$, que promove a morte do parasito pela ativação dos mecanismos microbicidas de macrófagos, associado ainda a produção de baixos níveis de IL-4 e IL-10 (CASTELLANO et al., 2009; SANTOS et al., 2013).

As células T $CD8^+$ são outro tipo de linfócitos envolvidos na resposta adaptativa. Essas células são capazes de reconhecer antígenos através da apresentação por MHC de classe I, sendo capazes de promover a morte de células infectadas através de mecanismos citotóxicos envolvendo a produção de granzimas e perforinas (MURPHY; WEAVER, 2017). Ainda que o controle da infecção seja feito majoritariamente através de $IFN-\gamma$ e células T $CD4^+$, linfócitos T $CD8^+$ antígeno-específicos também são ativados e expandidos no contexto da infecção por leishmania (BELKAID et al., 2002; SANTOS et al., 2013; UZONNA; JOYCE; SCOTT, 2004), e foram relacionadas com a cura clínica (DA-CRUZ et al., 1994, 2005). Esse efeito está relacionado principalmente à capacidade de produção de $IFN-\gamma$ por essas células, capaz de mediar a produção de IL-12 por células dendríticas, citocina necessária para a diferenciação T_H1 e para a ativação de macrófagos (KAYE; SCOTT, 2011; NOVAIS; SCOTT, 2015; SCOTT; NOVAIS, 2016).

Entretanto, embora as respostas T_H1 estejam associadas ao controle da infecção, os altos níveis de $TNF-\alpha$ e mediadores citotóxicos, típicos desse tipo de resposta, também têm sido associadas com formas mais graves da doença e cronicidade das lesões (BACELLAR et al., 2002; MELBY et al., 1994; SCOTT; NOVAIS, 2016). O estabelecimento da úlcera ocorre de forma conjunta ao aumento dos níveis de $IFN-\gamma$ e $TNF-\alpha$ e da diminuição dos níveis de IL-10, fatores que estão relacionados ao tamanho das lesões e à gravidade da doença (ANTONELLI et al., 2005; BACELLAR et al., 2002; CARVALHO, L. P. et al., 2012; FARIA et al., 2005; UNGER et al., 2009). Além disso, a menor expressão do receptor de IL-10 levando a uma diminuição no controle de mecanismos inflamatórios por essa citocina, também tem sido associada a imunopatogênese da leishmaniose cutânea (FARIA et al., 2005).

Neste sentido, apesar de células $CD8^+$ atuarem no controle da infecção e possuírem um papel protetor, elas também estão envolvidas com respostas imunopatogênicas e com o agravamento da doença levando à destruição tecidual (DA

SILVA SANTOS; BRODSKYN, 2014; SCOTT; NOVAIS, 2016). Foi demonstrado que essas células apresentam capacidade reduzida de produção IFN- γ no ambiente lesional da pele (NOVAIS et al., 2018). Ainda, o desenvolvimento de lesões ulceradas está associado com um aumento na frequência de células T CD8⁺ que expressam Granzima A (FARIA et al., 2009). A expressão de Granzima B e CD107a por essas células também foi correlacionada positivamente com o tamanho da lesão, evidenciando ainda mais o papel da atividade citotóxica na patogênese da lesão (SANTOS et al., 2013). Uma maior citotoxicidade também foi observada através do aumento na produção de Granzima B por células T CD8⁺ em pacientes, quando comparado a controles saudáveis ou pacientes com infecções subclínicas (CARDOSO et al., 2015). A estimulação de células mononucleares do sangue periférico (PBMC) *in vitro* com antígenos de *L. braziliensis* também demonstrou o aumento de células T CD8⁺ citotóxicas antígeno-específicas em um alto estado de ativação, em culturas celulares antes e durante o tratamento, bem como após cura clínica (CUNHA et al., 2020). De forma geral, os estudos sobre o papel dessas células durante a infecção apontam para um papel deletério das células T CD8⁺, evidenciando seus mecanismos citotóxicos como principais mediadores da patogênese.

1.4 A SENESCÊNCIA CELULAR

O fenômeno da senescência celular foi inicialmente identificado por Hayflick e Moorhead em 1961, que descobriram que fibroblastos apresentam uma capacidade finita de divisão celular em cultura (HAYFLICK; MOORHEAD, 1961). Esse fenômeno foi posteriormente chamado de senescência replicativa (HAYFLICK, 1965), e o limite de divisão celular ficou conhecido como Limite de Hayflick. Esse tipo de senescência é derivado da perda e disfunção dos telômeros após extensa atividade proliferativa na ausência da atividade da telomerase endógena (MUÑOZ-ESPÍN; SERRANO, 2014). Os telômeros são complexos formados por proteína e DNA que estão localizados nas extremidades cromossômicas em eucariotos, cuja função é proteger o material genético contra eventos que possam gerar instabilidade genômica (BLACKBURN, 2005). A erosão dos telômeros e a ausência de fatores compensatórios, como a telomerase, atingem um ponto crítico onde é reconhecida como uma quebra da fita dupla do DNA (AKBAR; HENSON, 2011; BLACKBURN, 2005), levando ao

recrutamento de mecanismos conhecidos como “resposta a danos no DNA” (DDR, do inglês: *DNA damage response*) (AKBAR; HENSON, 2011; CAMPISI; D’ADDA DI FAGAGNA, 2007; D’ADDA DI FAGAGNA, 2008).

O processo de senescência pode ocorrer tanto *in vivo* quanto *in vitro*, como resposta a um excesso de estresse extracelular ou intracelular, interrompendo o ciclo celular e impedindo que os danos sejam transmitidos para a próxima geração de células (CAMPISI; D’ADDA DI FAGAGNA, 2007; COPPÉ et al., 2010). Uma vez que a proliferação é interrompida, as células se tornam incapazes de duplicar seu material genético mesmo em condições adequadas de crescimento, um mecanismo que é causado pela expressão dominante de inibidores do ciclo celular (CAMPISI; D’ADDA DI FAGAGNA, 2007). Fatores que induzem dano grave ou irreparável ao DNA, deterioramento mitocondrial, estresse oxidativo, alterações na cromatina oriundas de estresse genotóxico e a expressão de certos oncogenes (senescência induzida por oncogenes) podem também estar associados à indução desse fenômeno, sendo regulados por vias de sinalização intracelular distintas, cuja principal função é regular o ciclo celular na tentativa de recuperar o dano sofrido (AKBAR; HENSON; LANNA, 2016; BALABAN; NEMOTO; FINKEL, 2005; BEN-PORATH; WEINBERG, 2004, 2005; BRAIG; SCHMITT, 2006; CAMPISI, 2005; CHIEN; KARSENTY, 2005; COPPÉ et al., 2010; D’ADDA DI FAGAGNA, 2008; DI LEONARDO et al., 1994; LOMBARD et al., 2005; MARTIN, 2005). A maior parte desses mecanismos é responsável pela indução de uma DDR semelhante àquela induzida pela erosão de telômeros, onde caso o dano ao DNA não seja reparado, as células entram em estado de interrupção da proliferação ou senescência (AKBAR; HENSON, 2011; AKBAR; HENSON; LANNA, 2016; D’ADDA DI FAGAGNA, 2008).

Assim, a senescência pode ser induzida por diversos mecanismos (RUFINI et al., 2013). O reconhecimento de DDRs envolve a ativação de quinases sensoras, como a ataxia-telangiectasia mutada (ATM) e a proteína relacionada a ATM e RAD3 (ATR, do inglês *ATM and RAD3 related*). Essa ativação resulta na formação de um foco de dano ao DNA, com o recrutamento de diversas proteínas, incluindo a histona H2AX fosforilada, para executar o reparo do DNA. A ativação de ATM e ATR resulta numa cascata de sinalização que leva à ativação de p53, uma proteína supressora de tumores, cuja via de sinalização leva ao estabelecimento da senescência (AKBAR; HENSON, 2011; CAMPISI, 2005). A fosforilação de p53 por ATM é um importante mecanismo de regulação da atividade de p53. Entretanto, essa proteína também

pode ser ativada por mecanismos independentes de ATM (LOMBARD et al., 2005). A proteína p53 possui papel crítico na indução desse fenômeno. Essa proteína ativa p21 e p16, proteínas inibidoras de ciclinas, que mediam a ativação da proteína Rb, que por sua vez interrompe a transcrição de genes e estabelece o processo de senescência (AKBAR; HENSON, 2011; BEN-PORATH; WEINBERG, 2005; CAMPISI, 2005; LOMBARD et al., 2005). Outra via de sinalização importante para o desenvolvimento desse processo é através da proteína quinase ativada por mitógeno p38 (p38 MAPK, do inglês *mitogen activated protein kinase*) (AKBAR; HENSON, 2011), que é ativado por estímulos dependentes ou não do telômero (IWASA; HAN; ISHIKAWA, 2003). A ativação de p38 MAPK é suficiente para induzir o processo de senescência através da ativação tanto de p53 quanto de p16 (FREUND et al., 2010).

O acúmulo de células senescentes em diversos tecidos está associado com o envelhecimento (CAMPISI et al., 2011). De forma interessante, fibroblastos que atingem o limite de Hayflick em cultura permanecem metabolicamente ativos e possuem algumas características marcantes. Essas células possuem como características associadas a senescência, o aumento na atividade da enzima β -galactosidase, aumento na expressão da proteína γ H2Ax (H2Ax fosforilada) associada ao complexo de reparo do DNA, e ativação constitutiva de proteínas p53, p38 MAPK e o inibidor de ciclina p16 (AKBAR; HENSON; LANNA, 2016; FREUND et al., 2010; ZGLINICKI et al., 2005). Essas células apresentam uma gama de alterações na organização da cromatina e expressão gênica levando a secreção de citocinas pró-inflamatórias, quimiocinas, fatores de crescimento e proteases caracterizando um fenômeno chamado de fenótipo secretor associado à senescência (SASP, do inglês: *senescence-associated secretory phenotype*) (CAMPISI, 2013; COPPÉ et al., 2008). Portanto, a senescência celular não representa apenas uma interrupção do crescimento, mas a obtenção de um fenótipo inflamatório característico (CAMPISI et al., 2011; FREUND et al., 2010).

Neste cenário, a aquisição de um perfil altamente inflamatório como um mecanismo compensatório aos defeitos funcionais replicativos e ao declínio de diversas outras funções fisiológicas está presente em diversos compartimentos celulares, incluindo o sistema imunológico. As mudanças que ocorrem nesse sistema associadas ao envelhecimento são conhecidas coletivamente como imunossenescência (GINALDI et al., 2001) e compreendem um conjunto complexo de fenômenos e alterações que ocorrem em vias intracelulares, moléculas de superfície

e metabolismo celular, caracterizando uma remodelagem das respostas imunológicas (FRANCESCHI; MONTI; BARBIERI; et al., 1995; FRANCESCHI; MONTI; SANSONI; et al., 1995; FRANCESCHI; BONAFÈ; VALENSIN, 2000; GINALDI et al., 2001). Com o envelhecimento, as respostas à antígenos previamente encontrados são conservadas, enquanto a habilidade de montar respostas efetivas contra novos antígenos diminui (FAGNONI et al., 2000), podendo resultar em alta susceptibilidade a doenças infecciosas e limitar a eficácia da vacinação de indivíduos idosos, contribuindo para morbidade e mortalidade (GINALDI et al., 2001). Isso ocorre devido a diminuição do número de linfócitos T *naïve* com o aumento da idade devido à involução do timo e com o aumento de linfócitos T de memória, os quais apresentam mudanças significativas em seu fenótipo e função, particularmente no compartimento de linfócitos T CD8⁺, culminando no acúmulo de células terminalmente diferenciadas (células T senescentes) (AKBAR; FLETCHER, 2005; LINTON; DORSHKIND, 2004; LYNCH et al., 2009; PEREIRA; AKBAR, 2016).

As células T senescentes compartilham diversas características com fibroblastos senescentes, como a redução na capacidade proliferativa, tamanho dos telômeros e liberação de citocinas próinflamatórias, além de elevada expressão de efetores citotóxicos como granzimas e perforinas (AKBAR; HENSON; LANNA, 2016), bem como a expressão constitutiva de vias associadas ao reparo de dano no DNA (LANNA et al., 2014). Além disso, os linfócitos T passam por diferentes estágios de diferenciação, evidenciados por diferentes marcadores de superfície celular em cada estágio (XU; LARBI, 2017). Células T em estágio senescente são predominantemente encontradas no compartimento de linfócitos T de memória efetora que reexpressam o marcador CD45RA (T_{EMRA}), caracterizadas pelo fenótipo CD45RA⁺CD45RO⁻CD27⁻CD28⁻CCR7⁻. Essas células estão associadas a perda da expressão de moléculas costimulatórias CD27, CD28 e do receptor CCR7 e aquisição de receptores de diferenciação celular como CD57 e KLRG1 (BRENCHLEY et al., 2003; EFFROS, 1997; HENSON; RIDDELL; AKBAR, 2012; PEREIRA; AKBAR, 2016; TARAZONA et al., 2001; WENG; AKBAR; GORONZY, 2009; XU; LARBI, 2017).

Apesar da reduzida capacidade proliferativa, linfócitos T senescentes permanecem metabolicamente ativos, sofrendo alteração na capacidade de produção e secreção de citocinas, quimiocinas e mediadores inflamatórios (como por exemplo: ROS, IL-1, IL-6, IL-8, TNF, GM-CSF, CXCL-1 e -2) (COPPÉ et al., 2008, 2010). Essa alteração, conhecida como SASP, é regulada através da fosforilação da MAP quinase

p38 (FREUND; PATIL; CAMPISI, 2011). Essa gama de alterações são caracterizadas por um estado pró-inflamatório crônico inespecífico, conhecido como *inflammaging* (*proinflammatory status of aging*) (FRANCESCHI et al., 2006), que já foi associado a diferentes comorbidades relacionadas ao envelhecimento (FREUND et al., 2010; FÜLÖP et al., 2016).

Além dessas características, os linfócitos senescentes T CD8⁺ e T CD4⁺ apresentaram ainda menor expressão de genes relacionados à via de ativação do receptor de células T (TCR) (LANNA et al., 2017; PEREIRA et al., 2020), potencialmente prejudicando a ativação e a resposta específicas desse grupo de células. Recentemente, nosso grupo demonstrou que células T_{EMRA} CD8⁺ adquirem um aumento na expressão de genes de citotoxicidade, com consequente aumento da função citotóxica em comparação com outras subpopulações de linfócitos. Além disso, essas células também adquirem maior expressão de genes de moléculas e receptores associados à células *natural killer* (NK), como a molécula adaptadora DAP-12 e os receptores da família KLR (do inglês, *Killer cell lectin-like receptors*), como KLRG1, NKG2C e NKG2D (PEREIRA et al., 2020). Essa alteração, com aumento da expressão de receptores de células NK e diminuição da atividade do TCR em células senescentes, é promovida pela expressão de proteínas sensores de estresse, denominadas Sestrin 2 (LANNA et al., 2017; PEREIRA et al., 2020). A expressão de receptores que normalmente são encontrados nas células NK por células senescentes foi proposta como um mecanismo de readaptação do sistema imunológico às alterações derivadas da senescência, como forma de manter a função celular mesmo que o reconhecimento antigênico esteja prejudicado (PEREIRA; AKBAR, 2016).

1.5 SENESCÊNCIA, INFECÇÕES E A PATOGÊNESE DA LCL

A inflamação crônica está associada ao envelhecimento natural e possui um papel importante em doenças relacionadas à idade, como câncer, aterosclerose e osteoartrite, diabetes e doenças neurodegenerativas (FREUND et al., 2010; FÜLÖP et al., 2016). Entretanto, além do envelhecimento fisiológico, estudos têm demonstrado o papel das infecções crônicas na indução de características de imunossenescência celular. A ativação crônica causada por infecções virais persistentes como o citomegalovírus (CMV) e o vírus Epstein-Barr (EBV) contribuem

diretamente para o acúmulo de células T CD8⁺ antígeno-específicas com características de senescência replicativa (OUYANG et al., 2003; PAWELEC; GOUTTEFANGEAS, 2006; PEREIRA; AKBAR, 2016). Em infecções causadas pelo vírus da imunodeficiência humana (HIV) também foram detectadas células T CD8⁺ antígeno-específicas com perda da capacidade de proliferação e redução do tamanho de seus telômeros (BRENCHLEY et al., 2003). O estado de estimulação antigênica persistente e o acúmulo de carga antigênica durante essas infecções causam a ativação progressiva de células do sistema imunológico, levando ao aumento da produção de fatores responsáveis pelo estado inflamatório (FREUND et al., 2010).

Recentemente, nosso grupo demonstrou de forma inédita que o estado de estimulação antigênica persistente, causado pela infecção humana por *L. braziliensis*, foi capaz de promover a senescência linfocitária prematura (COVRE et al., 2019). Essas células senescentes circulantes, foram identificadas no compartimento T_{EMRA}, e apresentaram características típicas de senescência celular, tanto em linfócitos T CD4⁺ quanto em T CD8⁺. As alterações observadas nos linfócitos durante a LCL incluíram: o aumento na expressão de CD57, KLRG1, γ H2Ax, fosforilação da proteína p38, bem como diminuição no tamanho dos telômeros e atividade do componente catalítico da telomerase. Os linfócitos senescentes também apresentaram ainda alta expressão de citocinas inflamatórias, mediadores de citotoxicidade e expressão de antígeno linfocitário cutâneo (CLA, do inglês: *cutaneous leucocyte-associated antigen*). Essas características evidenciaram uma alta capacidade funcional e potencial natural de migração para a pele dessas células que foram associadas ao dano tecidual e gravidade da lesão (COVRE et al., 2019).

As características de senescência celular descritas por nosso grupo durante a LCL, também foram observadas em células NK, que apresentaram um fenótipo de diferenciação tardia, evidenciado pelo aumento na expressão de CD57 e KLRG1, bem como pela diminuição da capacidade proliferativa e maior erosão telomérica. As células NK senescentes também apresentaram alta função efetora, caracterizada por uma elevada produção de citocinas inflamatórias e mediadores citotóxicos. Embora as células NK tenham demonstrado grande capacidade funcional, quando comparadas aos linfócitos TCD8⁺ senescentes, essas células apresentam baixa expressão de CLA e menor propensão de migração para a pele. Assim, dentre essas duas populações, os linfócitos T CD8⁺ senescentes parecem ser a principal população citotóxica mediando a destruição tecidual no local da lesão (COVRE et al., 2020).

Essas descrições demonstram a importância de se investigar a senescência neste contexto de doença, visto que as células senescentes foram capazes de mediar citotoxicidade e secretar mediadores inflamatórios, sugerindo que a atividade mediada por estas células podem ter um importante papel no desenvolvimento das lesões da LCL (COVRE et al., 2019; COVRE et al., 2020).

Cabe ainda ressaltar, no entanto, que os impactos da imunossenescência de células NK e T na LCL foram demonstrados utilizando majoritariamente células circulantes, o que nem sempre reflete a realidade do sítio lesional. Ainda, a expressão de genes relacionados a senescência não foi investigada no contexto da LCL, possibilitando ainda serem classificados como importantes alvos de investigação, que permitem avaliar o ambiente local da lesão de forma ampla. Assim, o estudo do transcriptoma se torna uma ferramenta de interesse para melhor compreensão dos fatores relacionados à senescência que influenciam a imunopatogênese dessa doença.

1.6 SEQUENCIAMENTO DE RNA E EXPRESSÃO GÊNICA

O desenvolvimento de um método de sequenciamento de ácidos nucleicos por Sanger, Nicklen e Coulson em 1977 (SANGER; NICKLEN; COULSON, 1977) revolucionou a ciência e o modo como os organismos vivos são estudados. O sequenciamento de Sanger, chamado de método de primeira geração, foi a metodologia de escolha por cerca de 30 anos devido as grandes vantagens que proporcionou sobre métodos anteriores (SCHUSTER, 2008). A junção de técnicas de sequenciamento com métodos computacionais da bioinformática permitiu o surgimento de novas abordagens, como o sequenciamento *shotgun* (ANDERSON, 1981; STADEN, 1979), capazes de sequenciar porções maiores dos genomas estudados. A partir desses avanços, em 2001 foram publicados dois rascunhos do genoma humano (LANDER et al., 2001; VENTER et al., 2001). Em 2004, o Consórcio Internacional de Sequenciamento do Genoma Humano (*International Human Genome Sequencing Consortium*), publicou a primeira versão completa do genoma humano (atualmente *National Center for Biotechnology Information (NCBI) build 36*), cujo custo de produção foi estimado em US\$ 300 milhões (ABDELLAH et al., 2004; METZKER, 2010). Um dos grandes problemas do sequenciamento gênico era o elevado custo de

execução relacionado aos projetos. No entanto, com o surgimento das técnicas de sequenciamento de próxima geração (NGS, do inglês: *next generation sequencing*) no início de 2008, esse custo vem diminuindo drasticamente tornando cada vez mais possível a execução de estudos de sequenciamento (Figura 1). Essas novas tecnologias empregam diversas estratégias que dependem de uma combinação de métodos de preparação, sequenciamento e alinhamento ou montagem do genoma, cujo maior avanço foi possibilitar um sequenciamento massivo, na ordem de 1 bilhão de leituras por corrida, com baixo custo (METZKER, 2010).

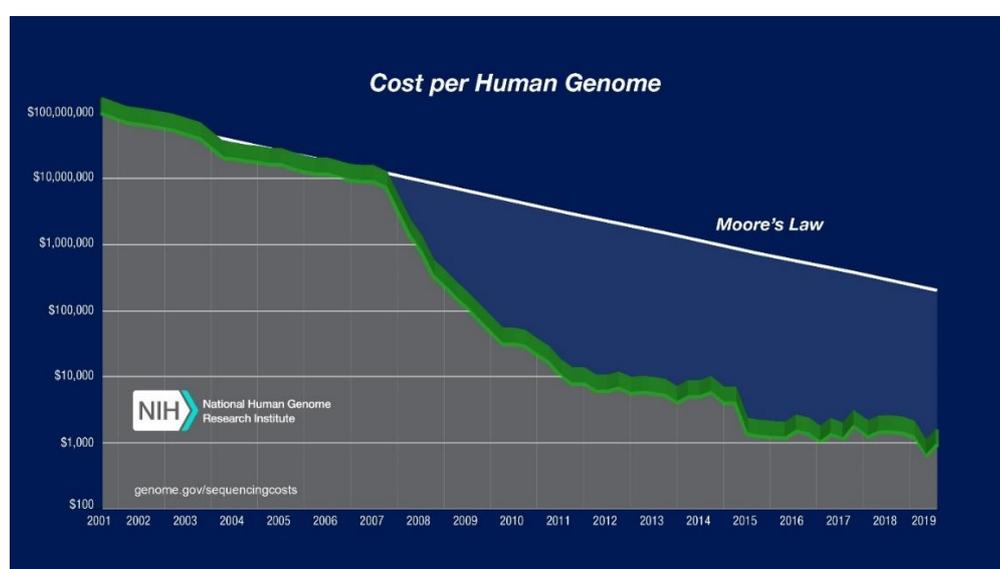


Figura 1 – Custo para o sequenciamento do genoma humano por ano. Nota-se um grande declínio a partir do início de 2008, onde houve a introdução das tecnologias de NGS. Fonte: <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>.

O transcriptoma é definido como o conjunto completo de transcritos, ou seja de todos os tipos de RNA, de uma célula bem como suas quantidades em um determinado estágio de desenvolvimento ou condição fisiológica (WANG; GERSTEIN; SNYDER, 2009). Medir a expressão gênica de um organismo em diferentes tecidos, condições ou pontos temporais é capaz de fornecer informações sobre como esses genes são regulados, revelando detalhes da biologia dos organismos, e tem sido instrumental no entendimento das doenças humanas (LOWE et al., 2017). Diversas tecnologias foram desenvolvidas com o objetivo de inferir e quantificar o transcriptoma, entre elas estão as metodologias que se baseiam na hibridização e aquelas baseadas no sequenciamento direto (WANG; GERSTEIN; SNYDER, 2009). Dentro essas metodologias destacam-se: a técnica de *microarrays*, baseada na

hibridização de sondas com moléculas alvo, tipicamente DNA complementar (cDNA), marcadas com fluoróforos, e; a técnica de sequenciamento de RNA (RNA-seq), baseado na determinação direta da sequência de cDNA de todos os RNA presentes na amostra de estudo (LOWE et al., 2017; MILLER; TANG, 2009; WANG; GERSTEIN; SNYDER, 2009). O RNA-Seq não depende do conhecimento prévio das sequências a serem determinadas, apresentando também menor custo e maior desempenho (WANG; GERSTEIN; SNYDER, 2009). Entretanto, a metodologia de RNA-seq produz conjuntos de dados grandes e complexos, que dependem criticamente de metodologias, que estão em constante evolução, para a análise e interpretação dos dados (OSHLACK; ROBINSON; YOUNG, 2010).

Uma análise típica de RNA-seq possui diversas etapas, que dependem fortemente da bioinformática. Após o sequenciamento dos fragmentos de cDNA, as leituras desses fragmentos são primeiramente mapeadas ao genoma ou transcriptoma, e são então agregadas ao nível de genes, éxons ou transcritos em uma matriz de expressão. Essa matriz é então normalizada e testada de acordo com o método estatístico de escolha para testagem diferencial de genes, gerando uma lista de genes diferencialmente expressos (GDEs) com valores de p e alteração em fator multiplicativo (*fold change*) associados, das quais são obtidas informações biológicas relevantes ao experimento (OSHLACK; ROBINSON; YOUNG, 2010). Assim, a análise de expressão diferencial se torna uma importante parte dessas análises, necessária para o entendimento e interpretação dos dados (LOVE; HUBER; ANDERS, 2014).

1.7 A EXPRESSÃO GÊNICA NA LCL

Poucos estudos têm analisado a expressão gênica durante as leishmanioses, especialmente no contexto da lesão. Durante a infecção por *L. braziliensis*, genes envolvidos com o recrutamento e ativação de linfócitos, células dendríticas (DCs), macrófagos, neutrófilos e células NK estão correlacionados com o controle da infecção e prevenção da evolução para formas mais graves da doença, como a LMC (MARETTI-MIRA et al., 2012). Genes que estão regulados negativamente na lesão incluem genes relacionados com a manutenção da barreira da pele, enquanto genes cuja expressão sofreram aumento são genes envolvidos com o recrutamento de células inflamatórias, com a atividade citotóxica e o inflamossoma (AMORIM et al.,

2019; NOVAIS et al., 2015). Genes inflamatórios e associados à resposta citotóxica estão correlacionados com a falha terapêutica durante a terapia com antimoniais pentavalentes, em particular os genes *GNLY* (Granulisina), *IL1B* (IL-1 β) e *IFNG* (IFN- γ) foram capazes de prever o desfecho de tratamento com acurácia de 86-96% (AMORIM et al., 2019). Além disso, a expressão de genes de citotoxicidade está correlacionada positivamente com a frequência de linfócitos T CD8⁺ e células NK (AMORIM et al., 2019), que são as células responsáveis por exercer esse mecanismo efetor. Genes inflamatórios, como IFN- γ e TNF- α , também estão associados com a infecção por *L. braziliensis*, assim como alguns genes inibitórios como IL-10, PD-1 e CTLA-4 (CHRISTENSEN et al., 2016). De forma complementar, a persistência desse parasito parece estar associada à uma resposta humoral, caracterizada pela expressão de genes de imunoglobulinas e genes relacionados à proliferação de linfócitos B (CHRISTENSEN et al., 2016). Entretanto, esses estudos transcriptômicos na LCL não avaliaram o impacto do fenômeno da senescência na imunopatogênese dessa doença.

2. JUSTIFICATIVA

As lesões nas leishmanioses são dependentes majoritariamente da combinação entre a resposta imunológica do hospedeiro e a espécie do parasito. Deste modo, respostas altamente inflamatórias estão associadas ao grau de gravidade da doença, observadas nas diferentes manifestações clínicas. A imunossenescência é um contexto imunológico em que a resposta celular é remodelada e cuja característica principal é a manutenção de um ambiente inflamatório crônico. Estudos recentes do nosso grupo demonstraram que células T senescentes acumulam-se durante a LCL e estão diretamente associadas à patogênese, através da grande capacidade de secreção de citocinas proinflamatórias e atividade citotóxicas.

Apesar de a expressão gênica ter sido investigada de forma geral no âmbito da LCL, nenhum estudo avaliou aspectos relacionados à imunossenescência no ambiente lesional. Uma vez que o papel da senescência vem sendo demonstrado no contexto da LCL, faz-se necessário uma caracterização mais ampla das células senescentes, mas também de quais funções inflamatórias elas exercem no sentido de aumentar a gravidade da doença. Neste contexto, o presente trabalho se propõe a analisar os perfis transcricionais com relação a genes sabidamente envolvidos na senescência no âmbito da lesão durante a infecção por *L. braziliensis*. Essa nova abordagem na análise da resposta local, levando em consideração as características imunológicas de senescência, é fundamental não somente para a compreensão dos mecanismos imunopatológicos da doença, mas também por trazer novas perspectivas para estudos em vacinas e esquemas terapêuticos mais eficientes e menos agressivos.

3. OBJETIVOS

3.1 OBJETIVO GERAL

Analisar o perfil de expressão de genes associados à senescência e à imunopatogênese da leishmaniose cutânea humana localizada causada por *L. braziliensis*, a partir de dados transcriptômicos de domínio público depositados no *Gene Expression Omnibus*.

3.2 OBJETIVOS ESPECÍFICOS

- Analisar a expressão diferencial geral de genes *in situ* da lesão;
- Analisar a expressão de genes específicos associados à senescência e à citotoxicidade;
- Realizar uma análise funcional dos genes diferencialmente expressos;
- Correlacionar a frequência de genes específicos com a frequência de subtipos celulares, prevista a partir do perfil total de genes, encontrados na lesão.

4. MATERIAIS E MÉTODOS

Este trabalho consiste em um estudo observacional de caso-controle, onde foi realizada a reanálise de dados de sequenciamento de RNA, em que são comparadas a expressão gênica *in situ* na pele de indivíduos saudáveis com a pele lesionada de indivíduos infectados por *Leishmania braziliensis*. O estudo compreende diversas análises de bioinformática, que foram executadas no Laboratório de Imunologia Celular e Molecular do Núcleo de Doenças Infecciosas da Universidade Federal do Espírito Santo. Por se tratar de uso de dados pré-existentes e de domínio público, as devidas considerações éticas já foram tomadas pelos autores originais do estudo.

4.1 OBTENÇÃO DOS DADOS

Os dados utilizados nesse trabalho consistem em amostras de sequenciamento de RNA de domínio público obtidos do repositório *online* do NCBI intitulado *Gene Expression Omnibus* (GEO). O GEO é um repositório para dados de expressão gênica de métodos de alta performance (*high throughput*) como o RNA-Seq e *microarrays* (BARRETT et al., 2013). Através do número de acesso (*accession number*): GSE127831, foram obtidos os metadados de cada amostra e a matriz de expressão gênica não normalizada necessária para as análises de expressão gênica diferencial.

4.2 DESCRIÇÃO RESUMIDA DO ESTUDO QUE DEU ORIGEM AOS DADOS ANALISADOS

Os dados obtidos são provenientes do estudo de Amorim e colaboradores (2019), os quais realizaram o sequenciamento de RNA de amostras de pele proveniente de indivíduos saudáveis (controles endêmicos, n = 7) e de pele de lesões de leishmaniose cutânea causada por *L. braziliensis* e determinada através de Reação em Cadeia da Polimerase (PCR) e teste intradérmico positivo com antígeno de *Leishmania* (n = 21). As amostras de pele foram coletadas utilizando *punch* de 4mm antes do início do tratamento, as quais foram homogeneizadas para extração de RNA (AMORIM et al., 2019).

4.3 SOFTWARES UTILIZADOS

Todas as análises no presente estudo foram efetuadas através do ambiente de computação estatística e gráfica R versão 4.0.0 (R CORE TEAM, 2020) através do programa RStudio versão 1.2.1335 (RSTUDIO TEAM, 2018), exceto quando especificado. Os pacotes utilizados são citados conforme segue a metodologia abaixo.

4.4 ANÁLISE DOS DADOS DE RNA-SEQ

Inicialmente, a matriz de expressão gênica não normalizada foi importada para o ambiente do R utilizando o pacote *readr* (WICKHAM; HESTER; FRANCOIS, 2018), que faz parte da coleção *tidyverse* para importação e preparação de dados (WICKHAM et al., 2019). Os metadados das amostras foram obtidos utilizando a função *getGEO()*, do pacote *GEOquery*, utilizando o número de acesso citado anteriormente (SEAN; MELTZER, 2007). Posteriormente, os dados fenotípicos foram extraídos utilizando a função *pData()* do pacote *Biobase*. A seguir, os dados foram preparados e posteriormente analisados para obtenção dos genes diferencialmente expressos entre a pele de indivíduos saudáveis e a pele da lesão de indivíduos infectados por *L. braziliensis*.

4.4.1 PREPARAÇÃO DOS DADOS

Inicialmente, os metadados foram filtrados com o auxílio de funções da coleção *tidyverse* (WICKHAM et al., 2019) e funções base do R, com a finalidade de obter uma tabela que contivesse apenas os dados relevantes para as análises subsequentes. A tabela fenotípica original continha dados a respeito da metodologia utilizada, endereço do local original do estudo, entre outros, que foram removidos para obter uma tabela com 10 variáveis (dados fenotípicos) a respeito de 28 observações, correspondente aos indivíduos do estudo.

Em seguida, foi construído o objeto *DESeqDataSet*, necessário para as análises de expressão diferencial. Posteriormente, os dados de expressão gênica propriamente ditos foram arredondados para números absolutos e pré-filtrados para remover genes com expressão baixa. Para isso foram removidos os genes em que a

somatória das linhas, ou seja, dos valores daquele determinado gene para todas as amostras, fosse inferior ou igual à 10 transcritos. Assim, o conjunto de dados ocupa menos espaço na memória do computador, facilitando o seu uso. Arredondar os valores é necessário pois a metodologia de análise diferencial empregada nesse estudo requer que os valores de expressão gênica sejam números inteiros.

4.4.2 ANÁLISE DE COMPONENTES PRINCIPAIS

Para avaliar a qualidade dos dados, foi realizada uma análise de componentes principais (PCA, do inglês *principal component analysis*). Para essa análise, foi utilizada a função `plotPCA()`, disponível no pacote `DESeq2` (LOVE; HUBER; ANDERS, 2014), sobre os dados normalizados através do método *Variance Stabilizing Transformation* (VST), também disponível no pacote `DESeq2` através da função `vst()`. Essa função transforma os dados para uma escala de \log_2 através do ajuste por um modelo estatístico (LOVE; HUBER; ANDERS, 2014). Essa transformação específica é utilizada para tornar os dados homocedásticos (com variância igual ou aproximada), uma vez que dados heteroscedásticos podem fazer com que os resultados sejam dominados por genes com alta variabilidade e que são altamente expressos no conjunto de dados, e evitar o espalhamento de dados pra genes com contagens baixas, onde interferência aleatória pode dominar os sinais com significados biológicos nas amostras numa transformação logarítmica comum (LOVE; HUBER; ANDERS, 2014).

4.4.3 ANÁLISE DE EXPRESSÃO GÊNICA DIFERENCIAL

Para a análise de expressão gênica diferencial, foi utilizada a função `DESeq()` do pacote `DESeq2` (LOVE; HUBER; ANDERS, 2014). A metodologia do `DESeq2` utiliza uma distribuição binomial negativa para estimar os genes diferencialmente expressos. Esse método foi escolhido por possuir maior sensibilidade em comparação com outros métodos, especialmente se tratando de mudanças pequenas na expressão gênica (LOVE; HUBER; ANDERS, 2014). Utilizou-se como fórmula para a análise os diferentes grupos com relação a origem da amostra: pele de indivíduos controle (codificado como *Healthy skin*) ou da lesão de LCL (codificado como *CL lesion*). Nessa

análise, foram considerados como genes diferencialmente expressos (GDEs) os genes que tivessem o valor de p ajustado pelo método Benjamini-Hochberg menor que 0,05. O método de ajuste do valor de p já é automaticamente implementado pela função *DESeq()*. Além disso, os GDEs são calculados após uma filtragem independente da função *DESeq()*, que efetua uma remoção adicional de genes com contagem baixa.

4.4.4 AGRUPAMENTO HIERÁRQUICO

Para o agrupamento hierárquico das amostras no estudo utilizou-se a função *hclust()* do pacote *stats*, um pacote básico já incluído no R, utilizando os parâmetros padrões, através do método de ligação completa. Esse cálculo foi feito sobre as distâncias euclidianas calculadas pela função *dist()*, também já incluída no R. O método de ligação completa é capaz de identificar agrupamentos semelhantes e mais homogêneos, baseando-se na dissimilaridade entre as amostras. A função *click_rotate()* do pacote *dendextend* (GALILI, 2015) foi utilizada para rodar os eixos do dendrograma e aprimorar a sua visualização. Os genes foram agrupados dentro da função que cria o mapa de calor, também através do cálculo da distância euclidiana entre cada gene selecionado.

4.4.5 ANÁLISE FUNCIONAL

Com a finalidade investigar os processos biológicos envolvidos na resposta imunológica à LCL, realizou-se uma análise funcional através da ontologia gênica. Para isso, foi utilizado um processo de análise chamado Sobre-representação (*Over-representation*) dos genes (BOYLE et al., 2004), através do pacote *clusterProfiler* (YU et al., 2012) e da função *enrichGO()*. Para isso, selecionamos os genes diferencialmente expressos e os testamos utilizando a base de dados *Gene Ontology* (GO), selecionando apenas as categorias envolvidas em Processos Biológicos.

4.4.6 DECONVOLUÇÃO DE CÉLULAS IMUNOLÓGICAS

A deconvolução de células é o processo de identificação, através de um algoritmo, dos subtipos celulares que se encontram presentes em uma determinada amostra de expressão gênica total, baseando-se na expressão gênica de células individuais estudadas anteriormente. Métodos computacionais de deconvolução permitem que as populações celulares envolvidas em um determinado contexto sejam investigadas em tecidos complexos baseando-se em dados transcriptômicos sem que haja a necessidade do sequenciamento individual dessas células. A composição de subtipos de células do sistema imunológico é uma característica chave que determina a resposta imunológica e é crucial para o entendimento das doenças. Para essa análise, foi utilizado o programa ImmQuant (FRISHBERG et al., 2016), uma ferramenta que permite analisar a alteração nos tipos celulares entre dois grupos de amostras. Nesse estudo, utilizamos a base de dados DMAP (NOVERSHTERN et al., 2011) para inferir as alterações celulares entre o grupo controle e os pacientes, utilizando os dados de expressão gênica normalizados pelo método VST disponível no pacote *DESeq2*.

4.4.7 CORRELAÇÃO ENTRE GENES E POPULAÇÕES CELULARES

Com a finalidade de avaliar a relação entre a presença de determinados subtipos celulares e a expressão de genes importantes na resposta à infecção, foi efetuada uma análise de correlação. Utilizou-se o método de Correlação de Postos de Spearman para calcular o coeficiente de correlação ρ entre a frequência relativa dos subtipos celulares e a expressão gênica normalizada, considerando significativas os resultados com valor de p menor que 0,05. As correlações foram calculadas entre todos os genes e tipos celulares selecionados, as quais foram utilizadas para construir uma matriz de correlação.

4.4.8 VISUALIZAÇÃO DOS RESULTADOS

Para visualização dos resultados foram utilizados os pacotes *ComplexHeatmap* (GU; EILS; SCHLESNER, 2016), para gerar os mapas de calor dos dados de

expressão gênica e do resultado da deconvolução de células. O pacote *ggplot2* (WICKHAM, 2016), também parte da coleção do *tidyverse*, foi utilizado para os demais resultados da expressão gênica e para a matriz de correlação. Para fins de visualização, os valores brutos de expressão gênica também foram normalizados utilizando a função *vst()*. Para visualização dos dados de ontologia, foram utilizadas as funções do pacote *enrichplot* (YU, 2019), implementado junto do pacote *clusterProfiler*.

4.5 SCRIPT E INFORMAÇÕES DE SESSÃO

Para fins de reprodutibilidade, o *script* contendo todo o código utilizado e as informações de sessão (dadas pela função *sessionInfo()* do R) estão disponibilizadas no Apêndice E.

5. RESULTADOS

5.1 EXPRESSÃO GÊNICA DIFERENCIAL

Foi efetuada uma análise de expressão gênica diferencial através do método do pacote *DESeq2*, o qual possibilitou a identificação de genes diferencialmente expressos entre a pele saudável e a pele lesional durante a infecção por *L. braziliensis*.

Originalmente, o conjunto de dados apresentava 34.935 genes totais. Após as filtrações necessárias, restaram 26,017 genes, os quais foram testados para expressão diferencial. No total foram identificados 12.306 GDEs, dos quais 6.076 foram regulados positivamente e 6.230 foram regulados negativamente (Tabela 1).

Tabela 1 – Descrição da análise de expressão gênica diferencial. * Estatisticamente significantes, valor de p ajustado < 0.05; ** % em relação ao número inicial de genes.

Análise de Expressão Gênica Diferencial		
	Número	Porcentagem**
Número inicial de genes	34.935	100,00%
Filtração inicial	8.390	24,02%
Total de genes analisados	26.545	75,98%
Filtração independente (DESeq2)	528	1,51%
Genes sem alteração	13.711	39,25%
Genes diferencialmente expressos*	12.306	35,23%
Expressão aumentada	6.076	17,39%
Expressão reduzida	6.230	17,83%

Além disso, uma análise dos GDEs demonstrou que embora o número de genes regulados negativamente e positivamente tenham sido semelhantes, houve uma maior amplitude de variação nos genes regulados positivamente. Além disso, a maior parte dos genes de interesse nesse estudo sofreram regulação positiva no contexto da infecção na pele (Figura 2).

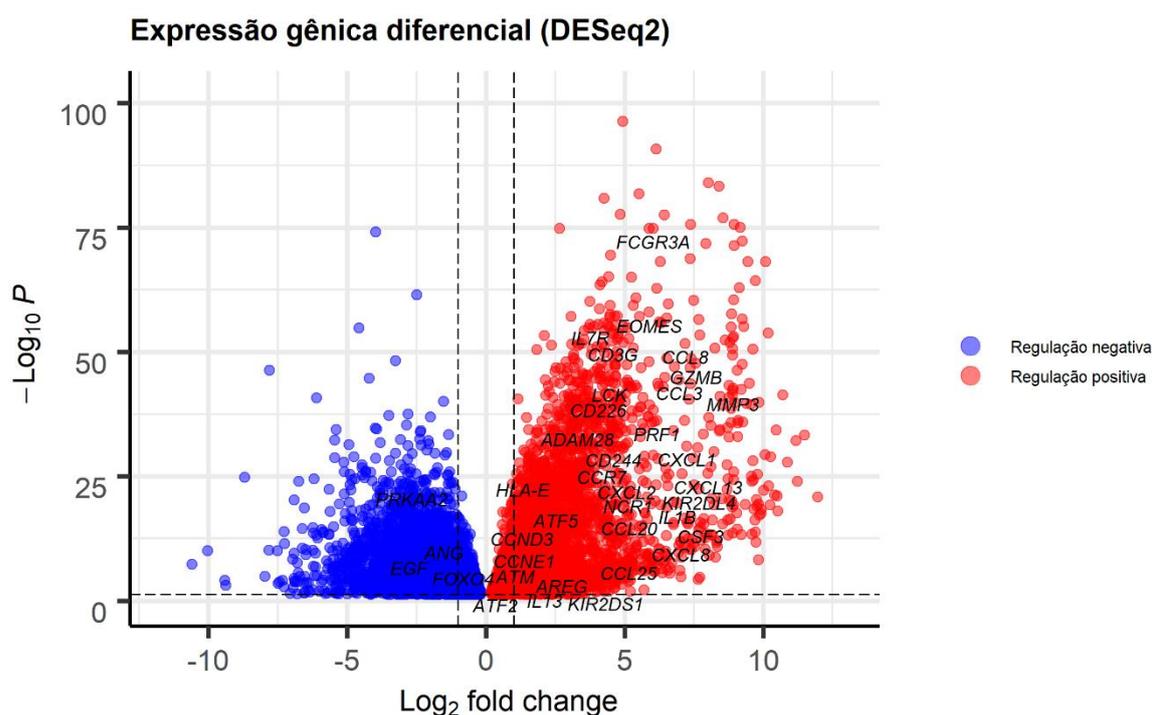


Figura 2 – Gráfico de vulcão dos genes diferencialmente expressos. O eixo y representa o valor de p, onde a linha tracejada horizontal equivale a um p ajustado de 0,05. O eixo x representa a alteração multiplicativa dos genes, em escala log₂, onde a linha tracejada vertical representa alteração equivalente à 1 (o dobro da expressão). Pontos azuis representam a regulação negativa, enquanto pontos vermelhos representam regulação positiva. Os genes identificados fazem parte dos genes de interesse analisados posteriormente nesse estudo.

5.2 ANÁLISE DE COMPONENTES PRINCIPAIS

Para avaliar a qualidade dos dados, foi efetuada uma análise de PCA dentro dos grupos estudados (Figura 3). Nessa análise, os componentes principais (PCs) 1 e 2 juntos foram capazes de representar a maior parte da variação dos dados (78%), indicando serem suficientes para distinguir os grupos amostrais. Ainda, o PC1 (eixo x) conteve a maior parte da variação, classificando-se como melhor componente para distinguir os grupos. Nesse resultado, observamos que a expressão dos genes de amostras correspondentes aos controles (CS, em azul) e pele lesional (LCL, em vermelho) agruparam-se em *clusters* espacialmente distintos. Isso demonstrou que a variação na expressão gênica presente nesse conjunto de dados foi capaz de distinguir os dois grupos dentro desse estudo de forma precisa.

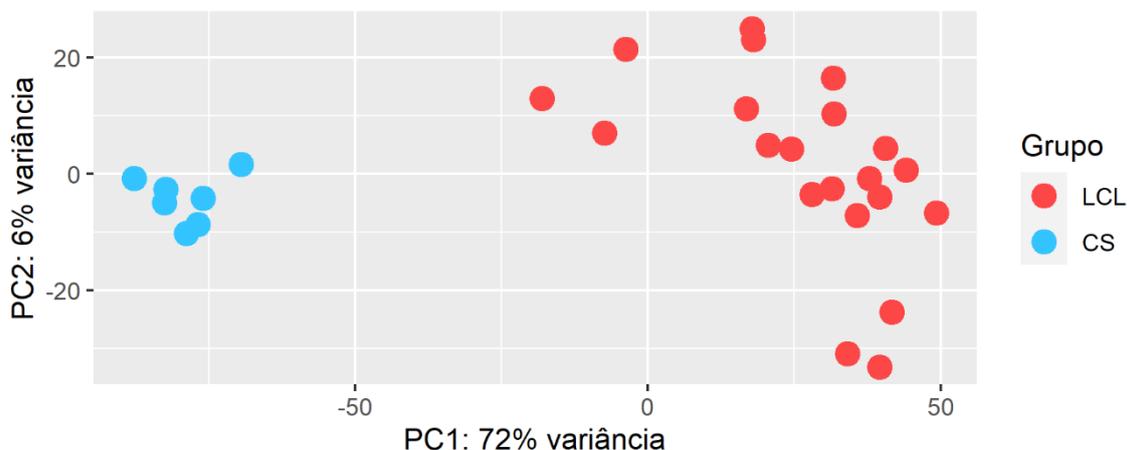


Figura 3 – Análise de Componentes Principais das amostras estudadas, contendo todos os genes analisados. Indivíduos saudáveis (CS, em azul, n = 7) estão representados à esquerda, enquanto indivíduos infectados (LCL, em vermelho, n = 21) estão representados à direita.

5.3 ANÁLISE FUNCIONAL

Para saber em quais processos biológicos estavam envolvidos os genes diferencialmente expressos nesse estudo, realizamos uma análise de ontologia gênica, onde os genes diferencialmente expressos foram submetidos à uma análise de sobre representação (*over-representation analysis*). Essa análise resultou em 794 processos biológicos sobre representados. Dessas categorias, as 30 mais significativas (baseado no valor de p) foram representadas (Figura 4), onde a maioria das categorias descritas estão relacionadas majoritariamente à processos biológicos envolvidos na resposta imunológica ou fenômenos do sistema imunológico. Dentre esses, foram destacados os processos de diferenciação de leucócitos, regulação de respostas efetoras, diferenciação de linfócitos, regulação positiva da ativação de leucócitos, diferenciação de células T, secreção de citocinas e regulação da resposta inflamatória. Além das categorias mostradas na figura, as categorias “envelhecimento” (*aging*) e “senescência replicativa” (*replicative senescente*) também estiveram presentes nesse resultado, com significância estatística (Apêndice A).

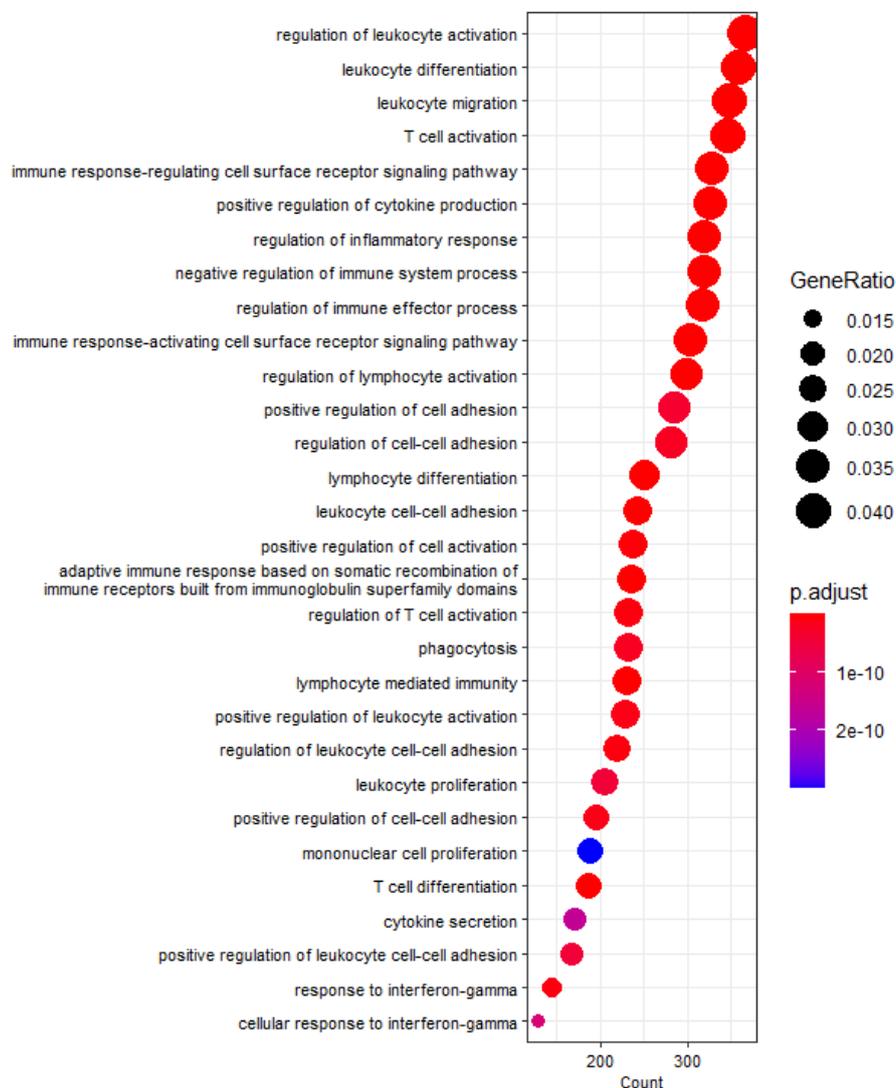


Figura 4 – Análise de sobre representação de ontologia gênica. No gráfico, o tamanho dos pontos reflete a proporção de genes em relação aos genes diferencialmente expressos analisados na ontologia, e o eixo x representa o número de genes sobre representados na categoria disposta no eixo y.

5.4 ANÁLISE DE GENES RELACIONADOS À SENESCÊNCIA

O papel da imunossenescência na infecção por *L. braziliensis* tem sido evidenciado por nosso grupo nos últimos anos, o qual demonstra exercer um impacto negativo na resolução da doença e está associado com a patogênese da LCL (COVRE et al., 2019, 2020). Deste modo, com o objetivo de corroborar esses resultados anteriores de nosso grupo (COVRE et al., 2019, 2020), selecionamos e categorizamos 179 genes associados a este processo. Esses genes foram diretamente relacionados

com características fenotípicas e funcionais associadas a células senescentes, além das características imunopatológicas da LCL. Desses genes, 143 genes foram diferencialmente expressos entre o grupo controle e o grupo de pacientes (Apêndice B).

Após a seleção dos genes relacionados à senescência que foram considerados diferencialmente expressos, esses genes foram utilizados para agrupar os indivíduos do estudo de acordo com sua expressão através do método de agrupamento hierárquico. Nesse resultado, o dendrograma demonstrou que os indivíduos foram separados em dois grupos principais distintos. Um composto majoritariamente por indivíduos controle saudáveis (codificado como HS1-HS7), à esquerda, e outro composto completamente por pacientes com LCL (codificado como CL1-CL21), à direita (Figura 5). Embora um dos pacientes tenha sido agrupado junto com os indivíduos controle (CL21), este foi identificado como indivíduo de maior dissimilaridade dentro do grupo (altura do dendrograma em aproximadamente 25). Esse mesmo agrupamento foi utilizado posteriormente para confecção dos mapas de calor.

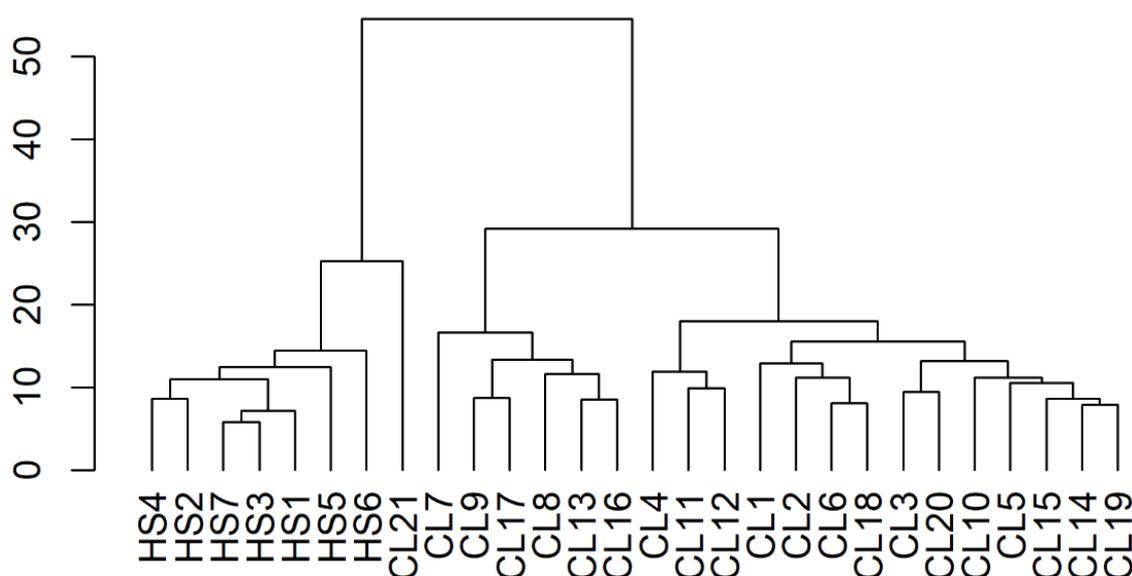


Figura 5 – Representação em dendrograma do agrupamento hierárquico das amostras do estudo. O agrupamento foi realizado através da diferença de expressão dos genes de interesse selecionados entre indivíduos controle (HS, n = 7) e pacientes com LCL (CL, n = 21). O eixo y representa a dissimilaridade entre os grupos, calculada através do método de ligação completa sobre as distâncias euclidianas entre as amostras.

Para melhor visualizar a alteração na expressão desses GDEs, foi construído um mapa de calor utilizando os genes de interesse agrupados pelas categorias pré-definidas, juntamente com o agrupamento hierárquico da figura anterior. Com exceção de um paciente no grupo LCL (CL21), cujo perfil foi semelhante ao dos controles para alguns genes, a expressão gênica foi distinta entre os grupos em todas as categorias de genes definidas (Figura 6). De forma geral, a maior parte dos genes selecionados teve sua expressão regulada positivamente no contexto da lesão cutânea durante a infecção por *L. braziliensis*. Deste modo, esse resultado indica a existência de uma assinatura transcricional capaz de distinguir entre o grupo controle e o grupo de pacientes infectados, baseada na expressão de genes de senescência na pele.

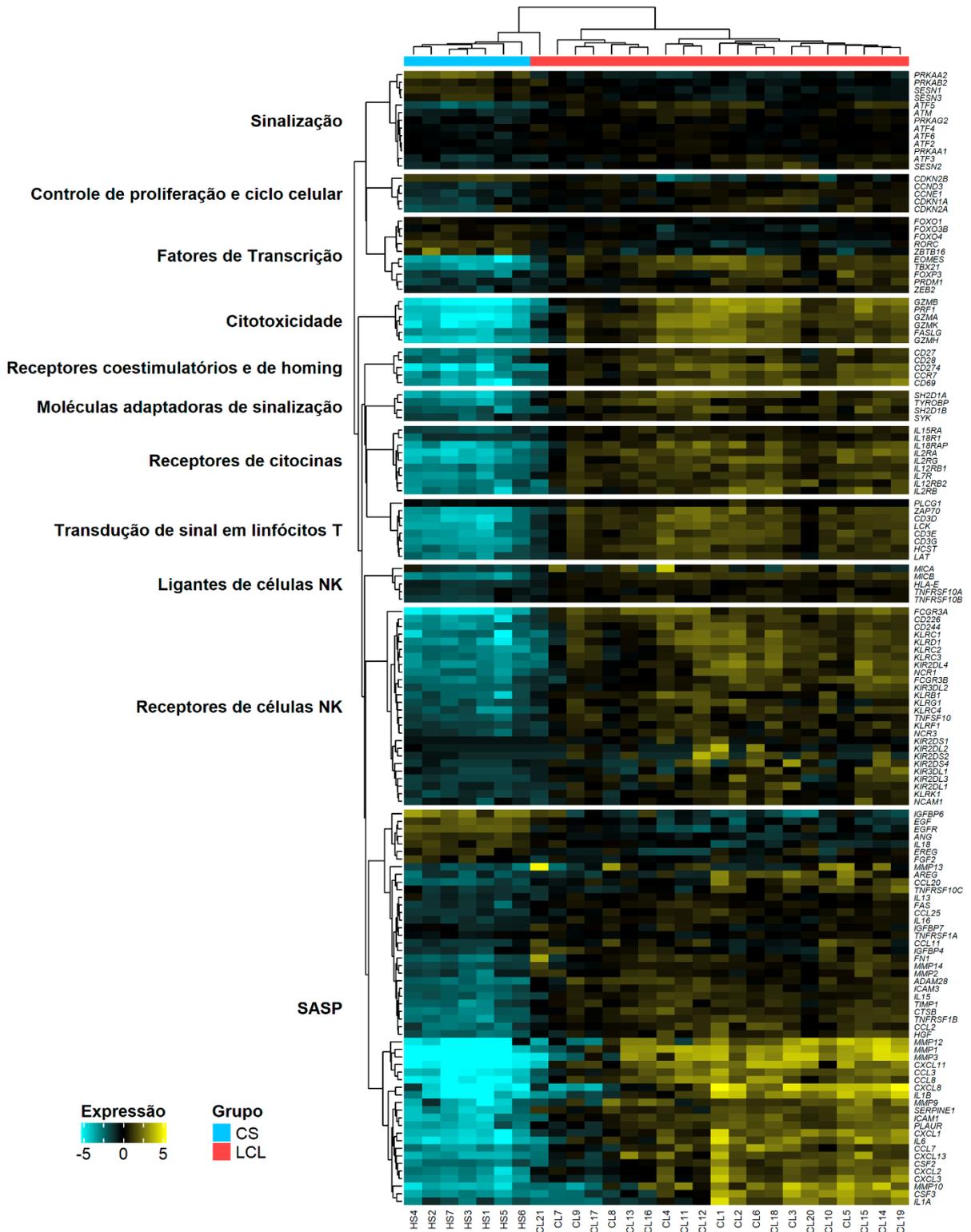


Figura 6 – Mapa de calor da expressão gênica, comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Os genes foram agrupados conforme as categorias predefinidas e explícitas no Apêndice B. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$.

5.4.1 MOLÉCULAS COESTIMULATÓRIAS E DE *HOMING*

Inicialmente, foram avaliados genes associados à resposta imunológica na LCL. Para isso, foi avaliado o ambiente de estimulação e endereçamento das células envolvidas na resposta imunológica, através da seleção de algumas moléculas. Estas incluíram, dentre os genes diferencialmente expressos, moléculas coestimuladoras (CD27, CD28 e CD69), inibitórias (PD-L1) e de *homing* (CCR7). Os resultados demonstram a ocorrência de um aumento na expressão de todos os genes citados (Figura 7).

5.4.2 RECEPTORES DE CITOCINAS

Foram selecionados genes de receptores de citocinas envolvidas na estimulação e proliferação de células T e células NK, importantes na resposta imunológica durante a LCL e potencialmente associados a mecanismos de imunorregulação dessa doença. Além disso, alguns dos genes selecionados são receptores de citocinas envolvidas no SASP. Todos os receptores de citocinas analisados tiveram sua expressão aumentada na pele durante a infecção por *L. braziliensis* (Figura 8).

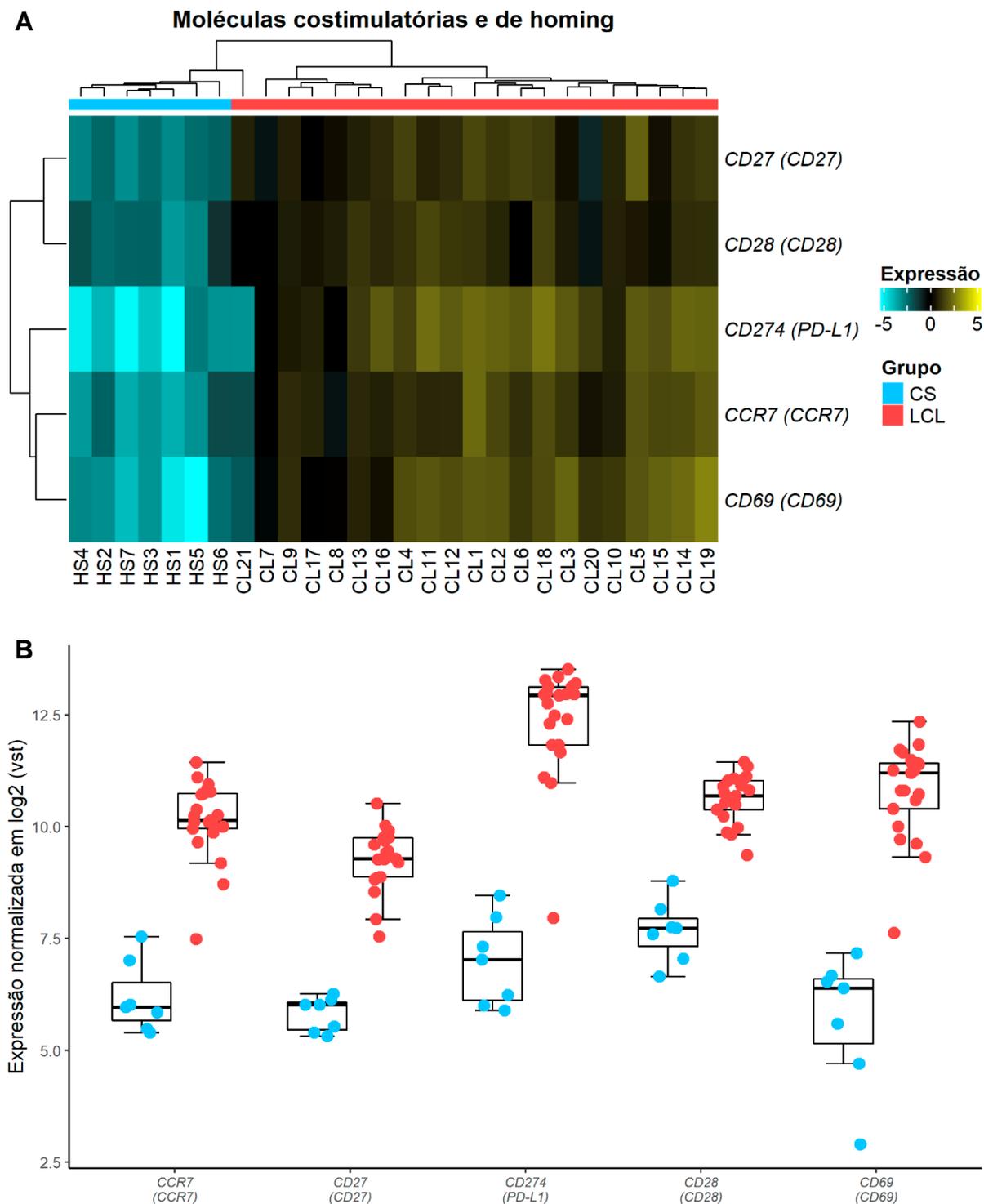


Figura 7 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de moléculas costimulatórias e de *homing* comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de < 0,05.

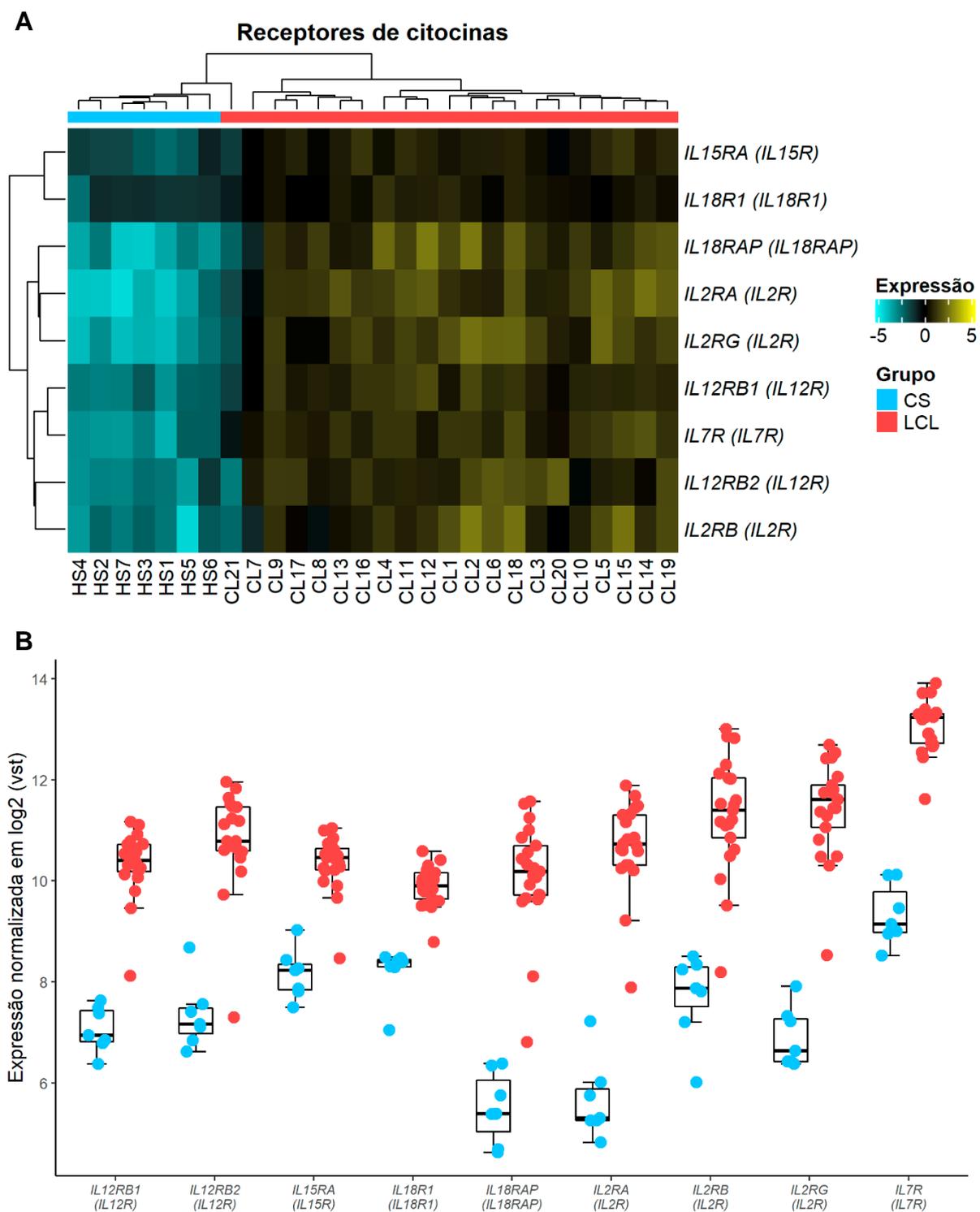


Figura 8 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de receptores de citocinas comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de < 0,05.

5.4.3 MOLÉCULAS ADAPTADORAS DE SINALIZAÇÃO

Ainda com a finalidade de avaliar a ativação das células envolvidas na resposta imunológica à *L. braziliensis*, foram selecionados genes codificantes para moléculas adaptadoras de sinalização. Nessa análise, foram avaliadas proteínas sinalizadoras para receptores de células NK e de células T, onde todos os genes analisados tiveram sua expressão regulada positivamente na lesão (Figura 9).

5.4.4 RECEPTORES DE CÉLULAS NK

Para investigar a associação de receptores de células NK com a resposta à *L. braziliensis*, foram selecionados 33 genes desses receptores para análise, incluindo o receptor CD56 e os receptores KLR e KIR (*killer-cell immunoglobulin-like receptor*). Desses genes, apenas sete não foram diferencialmente expressos. Dos genes diferencialmente expressos, todos tiveram sua expressão aumentada (Figura 10), demonstrando, portanto, um aumento na presença desses receptores na lesão. O gene *B3GAT1* (CD57), que codifica um marcador de senescência celular, não apresentou diferença entre os grupos (Apêndice B).

5.4.5 CONTROLE DE PROLIFERAÇÃO E CICLO CELULAR

Visando avaliar o crescimento e proliferação celular, foram selecionados genes envolvidos no controle do ciclo celular e da proliferação. Nossos resultados demonstraram que na lesão houve redução na expressão apenas do gene codificante da proteína p15 (gene *CDNK2B*), com aumento nos genes de ciclinas (*CCND3* e *CCNE1*) e também das proteínas p21 (*CDKN1A*), p16 (*CDKN2A*) e p15 (*CDNK2B*), indicando a presença de mecanismos de inibição da proliferação celular (Figura 11).

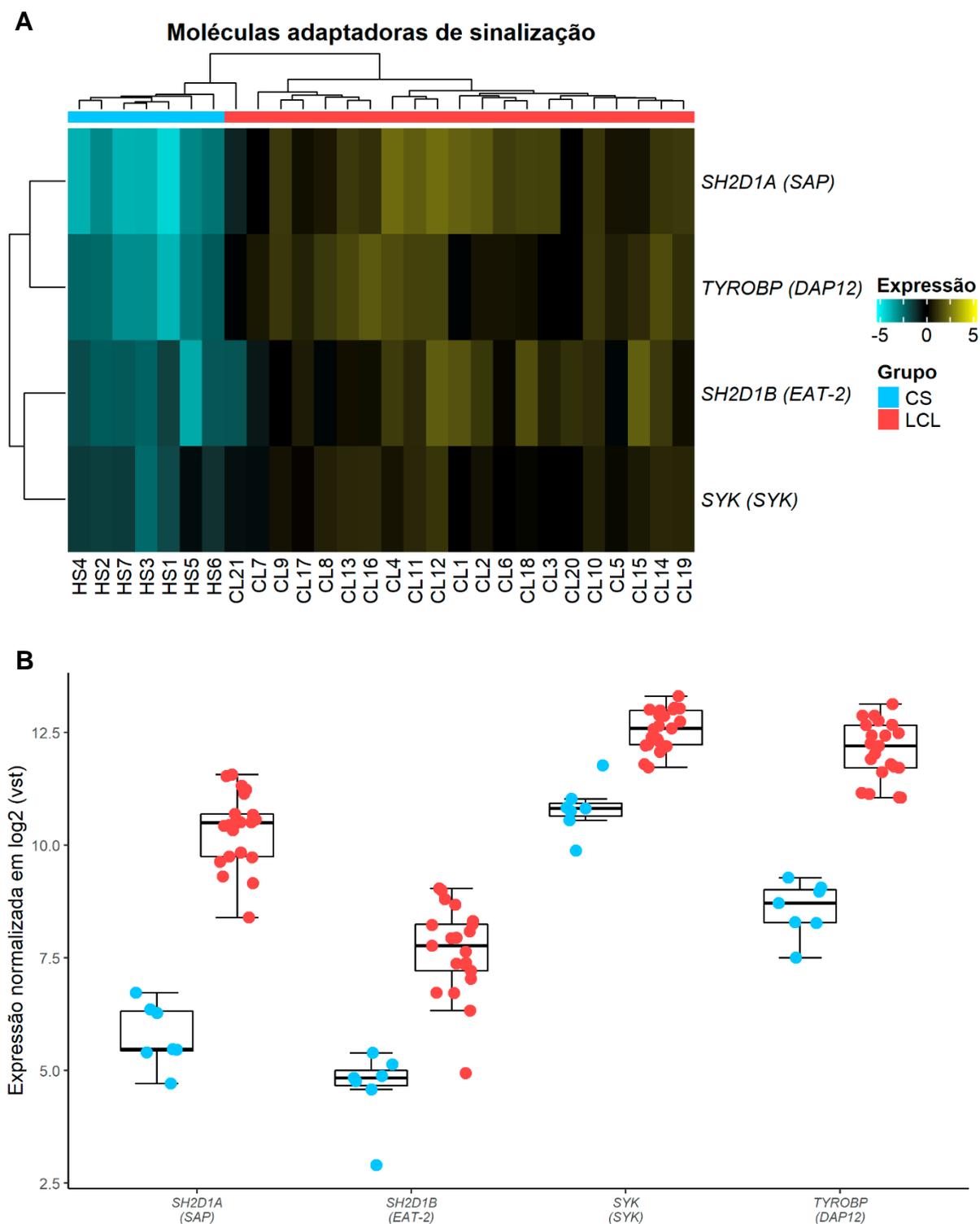


Figura 9 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de moléculas adaptadoras de sinalização (p ajustado $< 0,05$) comparando os grupos controle (CS, em azul, $n = 7$) com os pacientes (LCL, em vermelho, $n = 21$). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$.

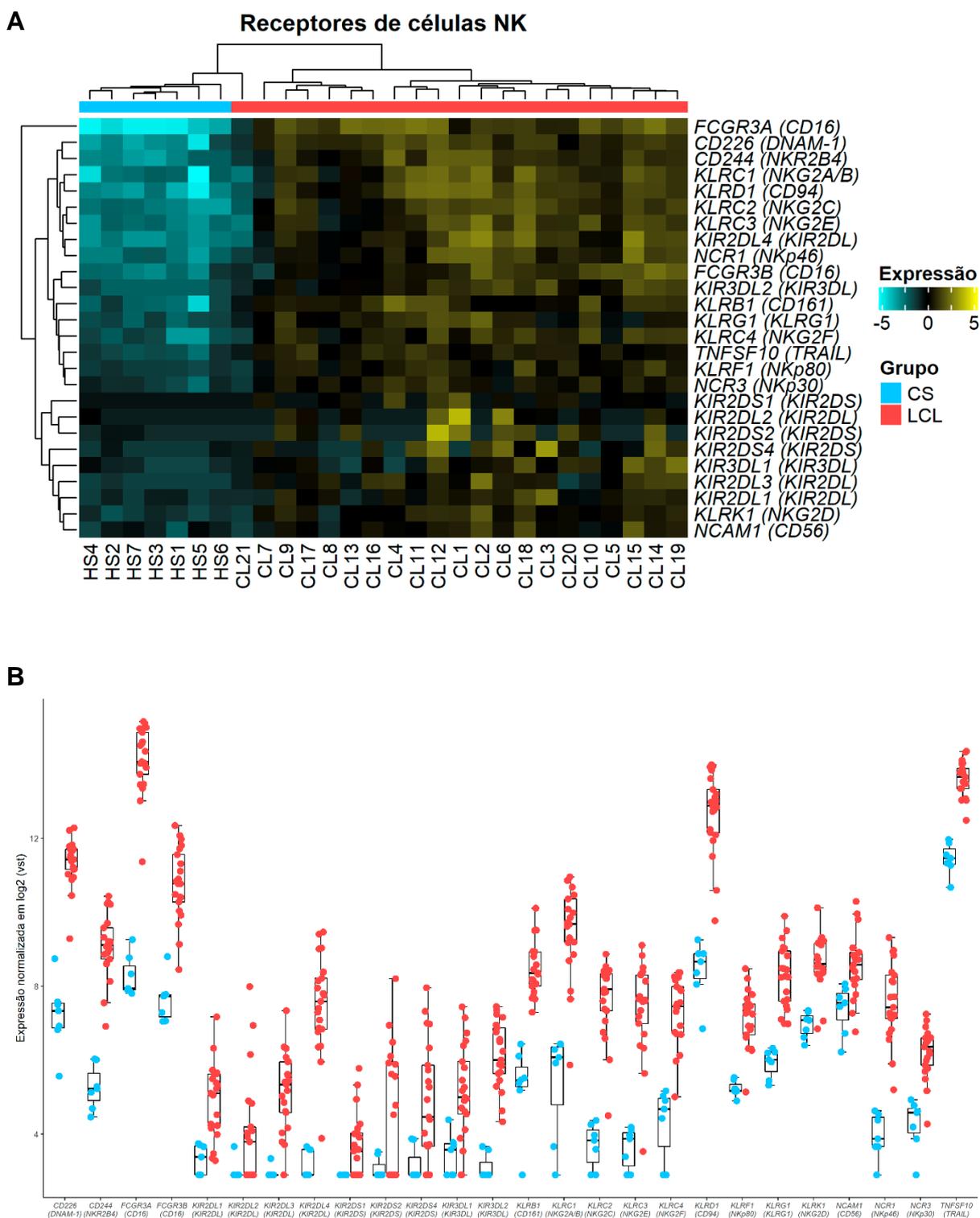


Figura 10 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de receptores de células NK comparando os grupos controle (CS, em azul, $n = 7$) com os pacientes (LCL, em vermelho, $n = 21$). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$.

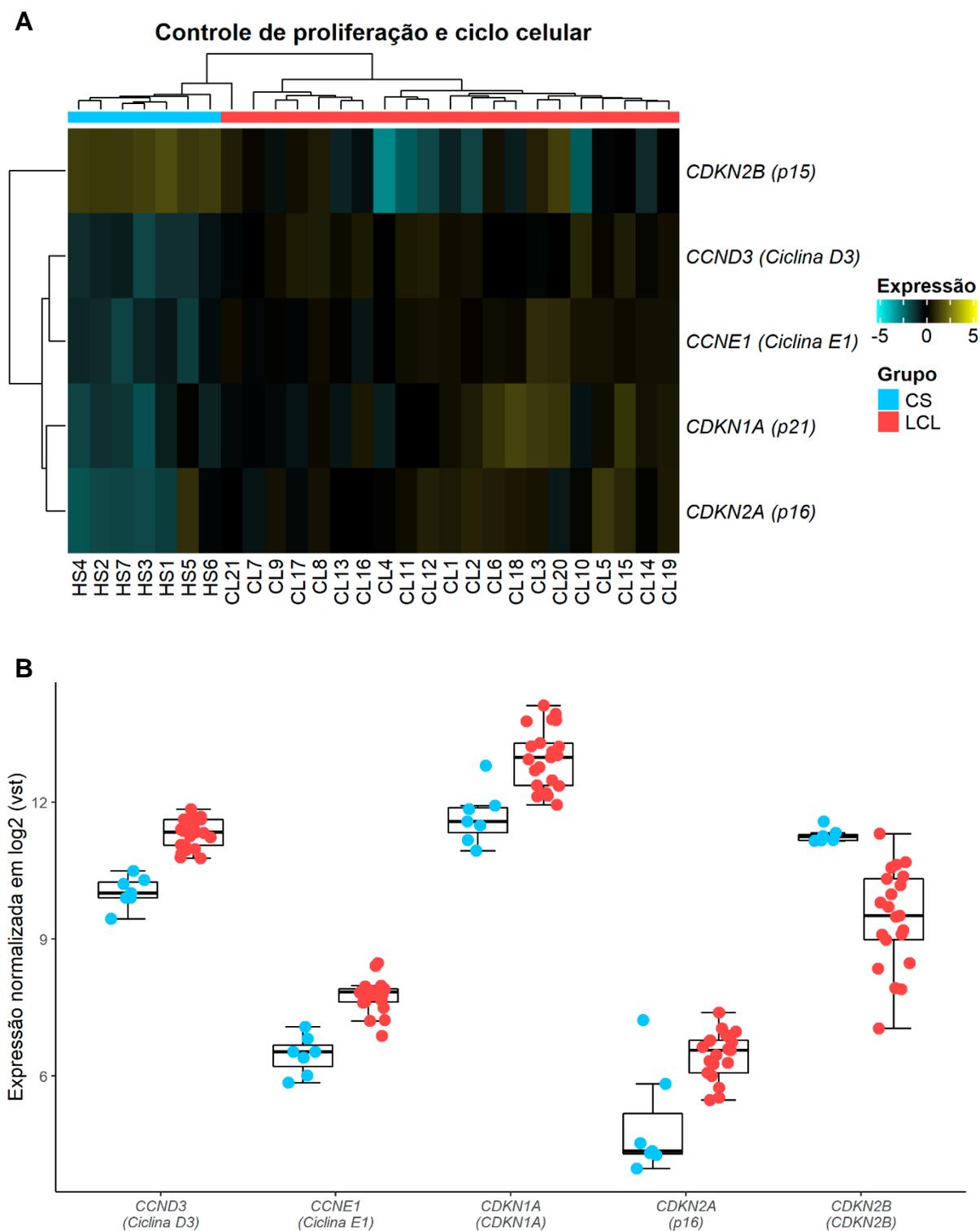


Figura 11 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de controle de proliferação e ciclo celular comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de < 0,05.

5.4.6 SINALIZAÇÃO

Foram avaliadas também moléculas envolvidas na sinalização celular potencialmente associadas à senescência, principalmente em resposta ao estresse celular. Nesse contexto, foram selecionados genes que incluem os fatores de transcrição de ativação (ATF, do inglês *activating transcription factors*), proteínas quinases fosforiladas ativadas por AMP (p-AMPK, do inglês *phosphorylated AMP-activated protein kinase*) e os sestrins, que estão diferencialmente expressos na lesão (Figura 12). Os resultados demonstram que os genes *PRKAA1* e *PRKAG2* estão aumentados na lesão, juntamente do gene *ATM*, dos genes para ATF e do gene *SESN2*, que codifica Sestrin 2. Em contraste, os genes para Sestrins 1 e 3 sofreram diminuição, juntamente dos genes *PRKAA2* e *PRKAB2*.

5.4.7 FATORES DE TRANSCRIÇÃO

Também para avaliar a proliferação e diferenciação celular, foram selecionados genes de fatores de transcrição envolvidos nesses processos. Os genes FOXO (*forkhead box transcription factors Class O*) estão envolvidos, entre outros mecanismos, na proliferação e diferenciação celular. Nossos resultados demonstraram uma redução na expressão desses genes, assim como os genes *RORC*, de diferenciação Th17, e o gene *ZBTB16*, de diferenciação de linfócitos NKT. Contrário a isso, os genes *EOMES* e *TBX21*, relacionados a maturação e diferenciação de células efetoras citotóxicas, apresentaram aumento. Em conjunto, *ZEB2* e *PRDM1*, que estão relacionados a diferenciação terminal de células T efetoras e o fator de transcrição *FOXP3*, necessário para o desenvolvimento de células T reguladoras, também foram regulados positivamente na lesão (Figura 13).

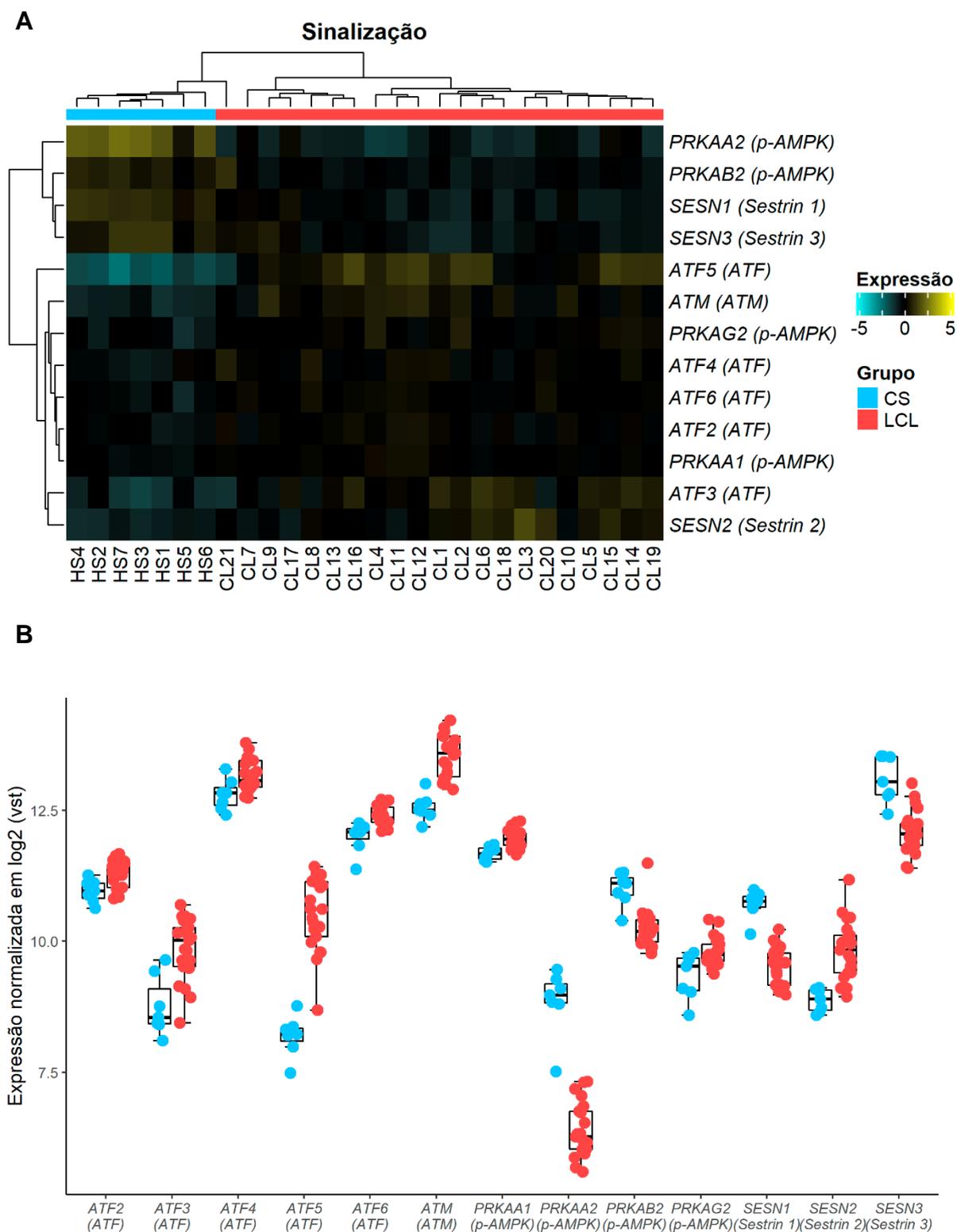


Figura 12 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de sinalização comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de < 0,05.

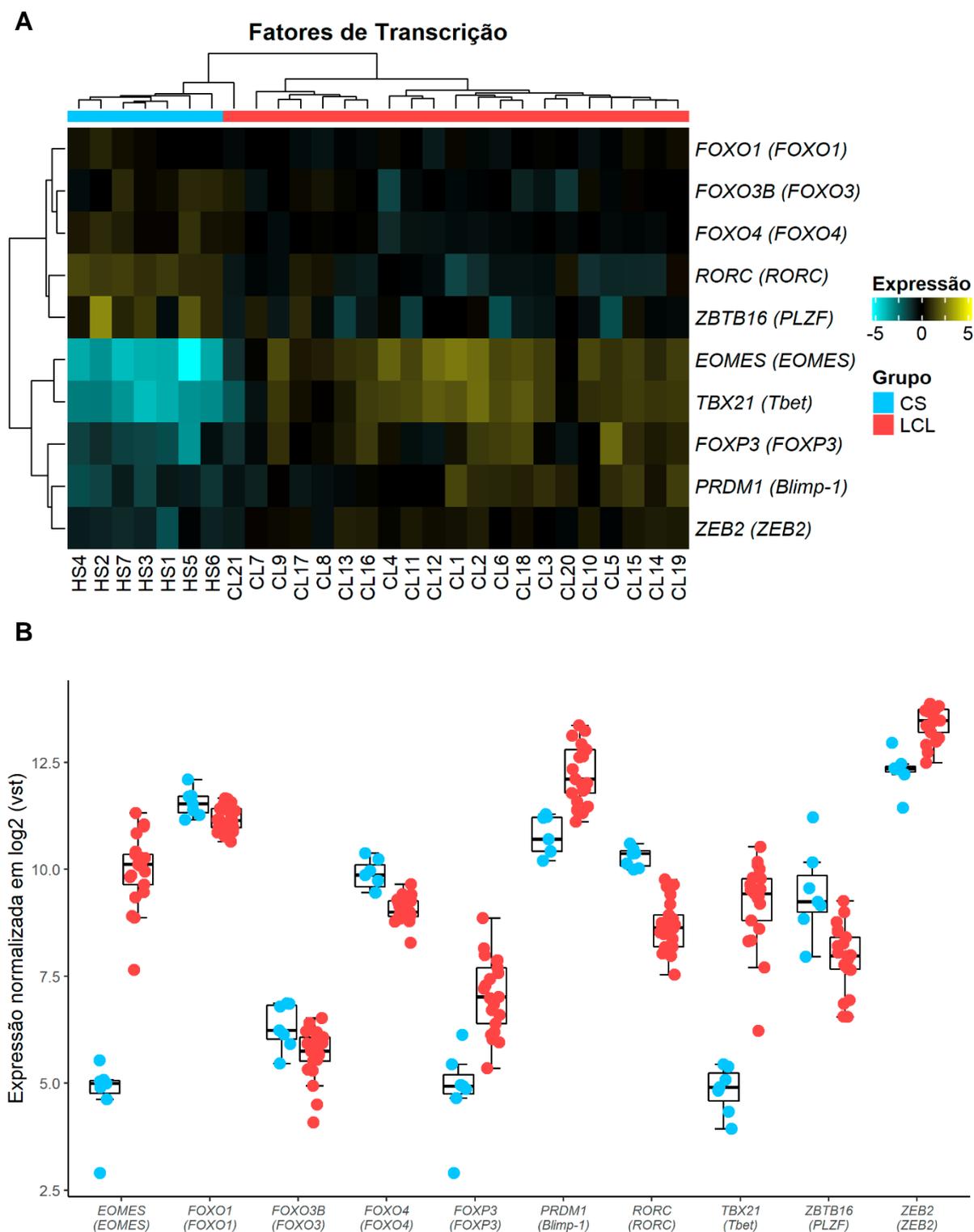


Figura 13 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de fatores de transcrição comparando os grupos controle (CS, em azul, $n = 7$) com os pacientes (LCL, em vermelho, $n = 21$). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$.

5.4.8 TRANSDUÇÃO DE SINAL EM LINFÓCITOS T

O receptor de células T (TCR, do inglês *T cell receptor*), em associação com o complexo de sinalização CD3 é o principal determinante do desenvolvimento e ativação das células T nas respostas à antígenos (DONG et al., 2019). Na presente análise, no grupo de genes “transdução de sinal em linfócitos T”, genes responsáveis pelas subunidades do complexo CD3 todos tiveram sua expressão regulada positivamente no contexto da lesão (Figura 14).

5.4.9 CITOTOXICIDADE

Para avaliar a ativação de mecanismos de citotoxicidade no contexto da infecção, selecionamos genes que codificam as proteínas envolvidas nesses mecanismos. Todos os genes analisados foram diferencialmente expressos e apresentaram um aumento considerável (Figura 15), indicando uma elevada presença desses mecanismos na pele lesionada.

5.4.10 LIGANTES DE CÉLULAS NK

Dos ligantes de células NK analisados, exceto pela E-Caderina, todos foram diferencialmente expressos e aumentados na lesão (Figura 16), incluindo os genes MICA e MICB, que ligam-se a NKG2D, cuja expressão se encontra aumentada na lesão (gene *KLRK1*, Figura 10), e os genes DR4 e DR5 (genes *TNFRSF10A* e *TNFRSF10B*) que se ligam ao receptor TRAIL, também aumentado na lesão (gene *TNFSF10*, Figura 10).

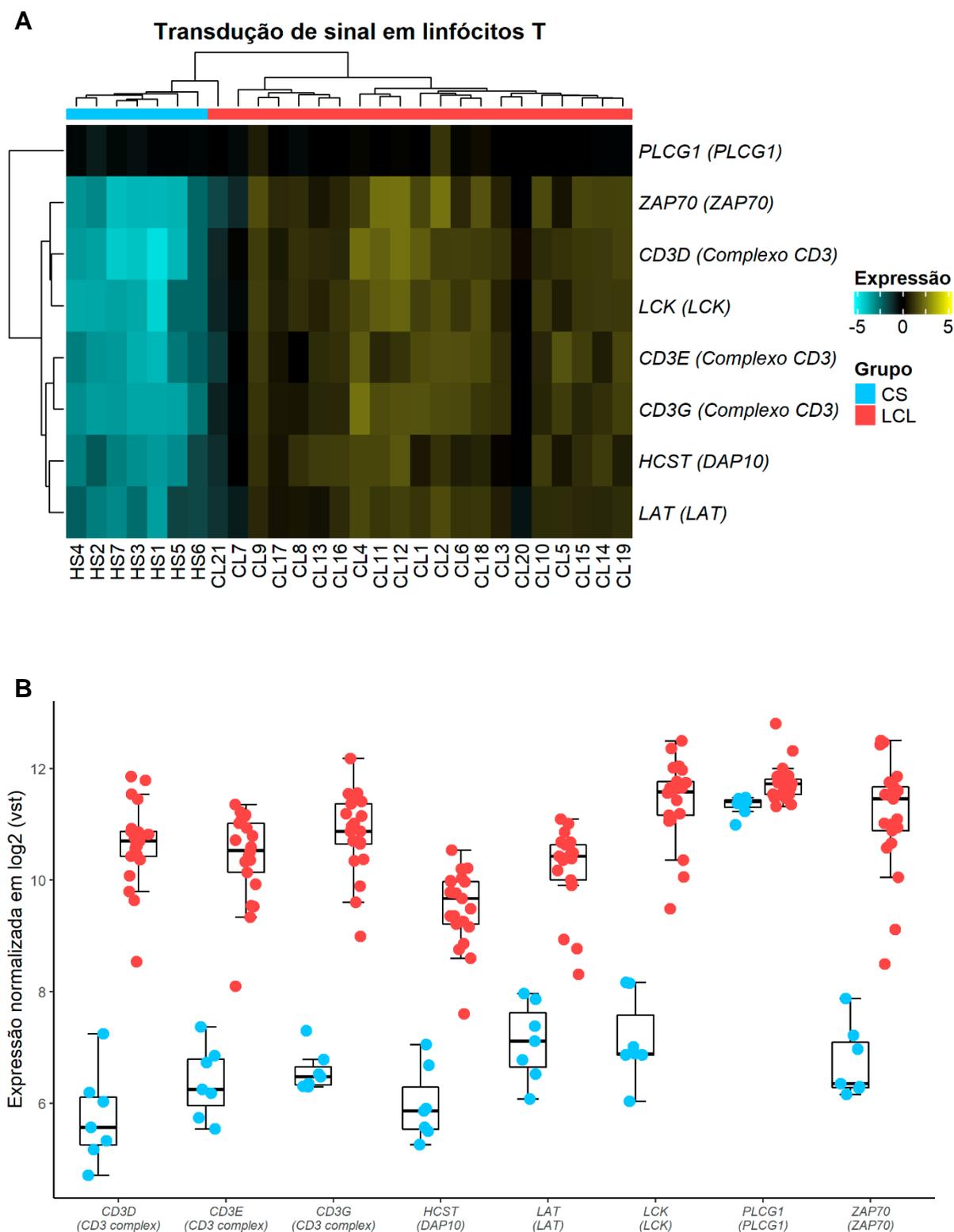


Figura 14 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de transdução de sinal em linfócitos T comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de < 0,05.

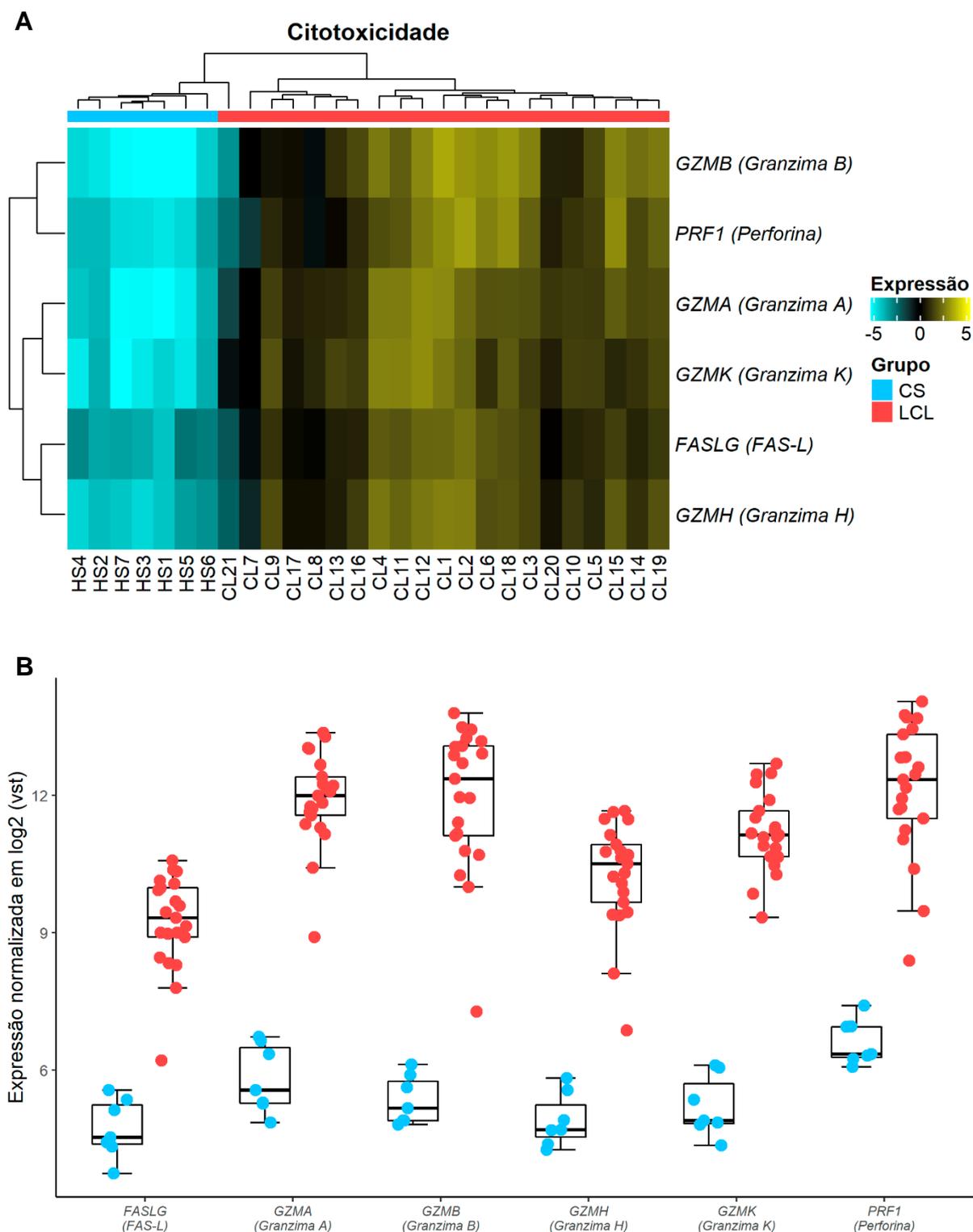


Figura 15 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de citotoxicidade comparando os grupos controle (CS, em azul, $n = 7$) com os pacientes (LCL, em vermelho, $n = 21$). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$.

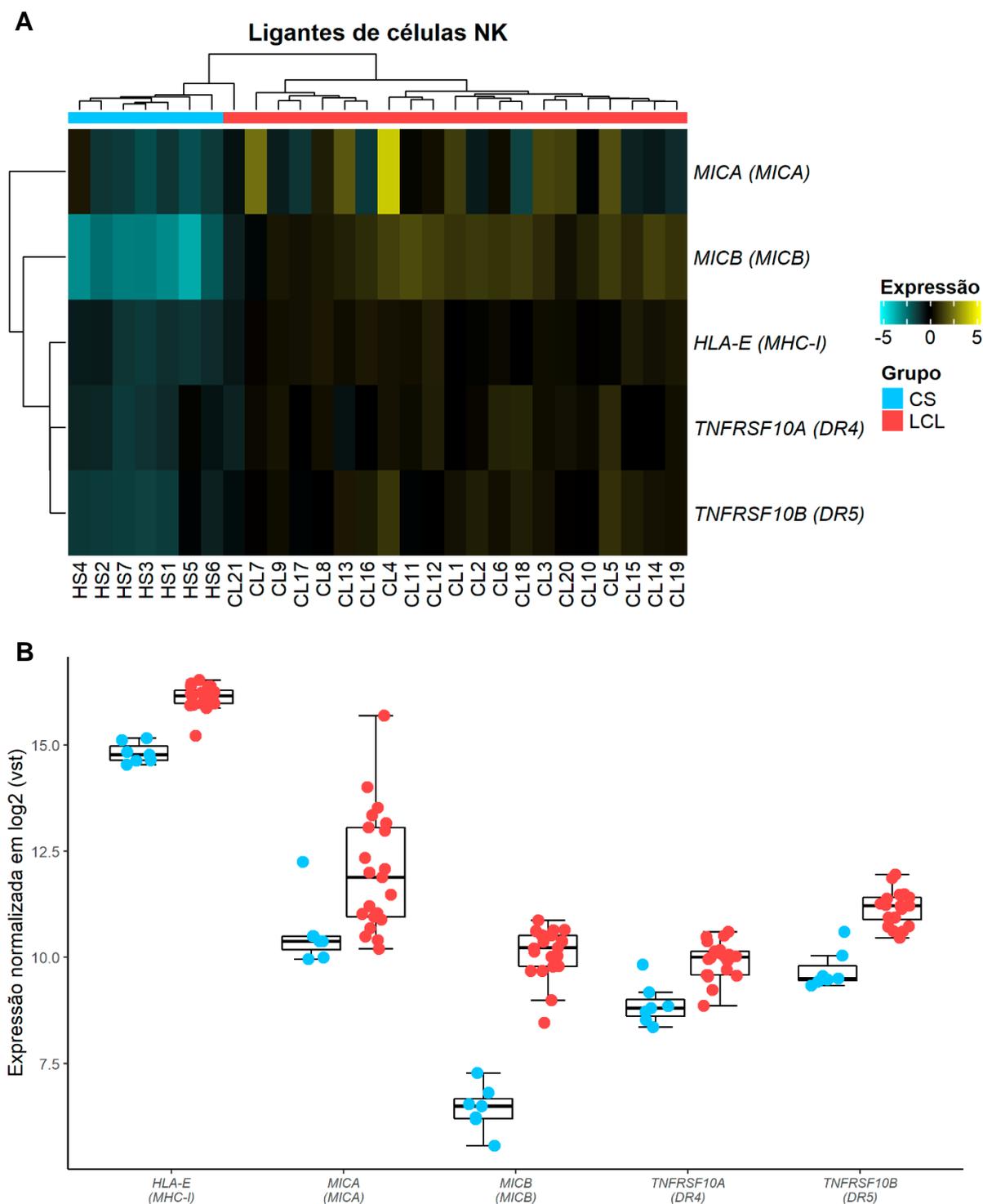


Figura 16 – Representação em mapa de calor (A) e diagrama de caixa com pontos (B) dos genes de ligantes de células NK comparando os grupos controle (CS, em azul, $n = 7$) com os pacientes (LCL, em vermelho, $n = 21$). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de $< 0,05$.

5.4.11 SASP

Observa-se que a maior parte dos genes diferencialmente expressos relacionados ao SASP teve sua expressão aumentada. Dos 71 genes analisados, apenas 19 não apresentaram diferenças entre os grupos. Dos genes diferencialmente expressos, apenas os genes *FGF2*, *EREG*, *IL18*, *ANG*, *EGFR*, *EGF* e *IGFBP6* tiveram sua expressão reduzida (Figuras 17 e 18). Esses resultados indicam que o perfil de secreção do SASP está presente no ambiente de resposta imunológica da lesão cutânea durante a LCL.

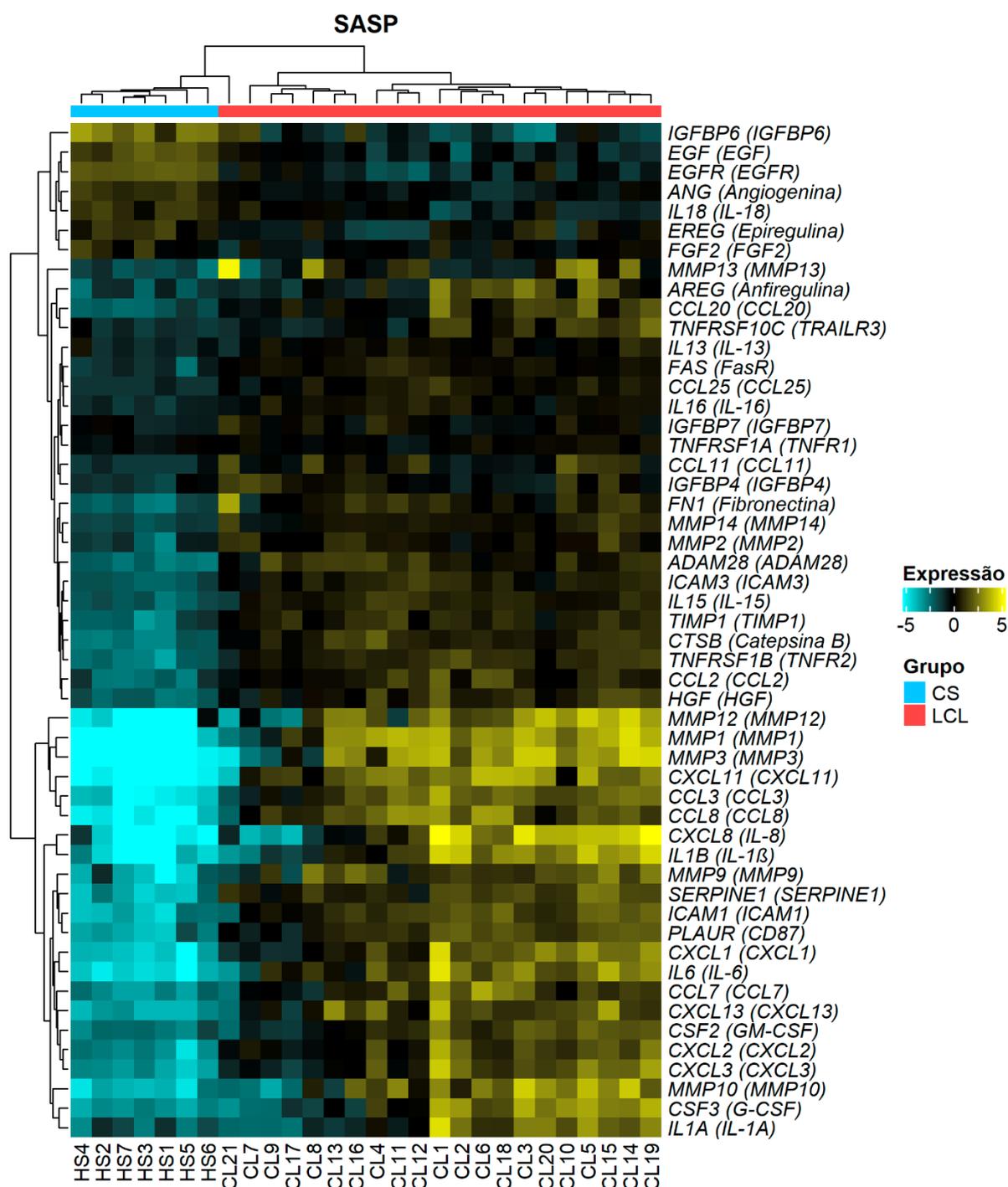


Figura 17 – Representação em mapa de calor dos genes pertencentes ao SASP comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). No mapa de calor, os valores estão centrados na média dos genes, com valores em azul ciano representando genes regulados negativamente, enquanto o amarelo representa os genes regulados positivamente. Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de < 0,05.

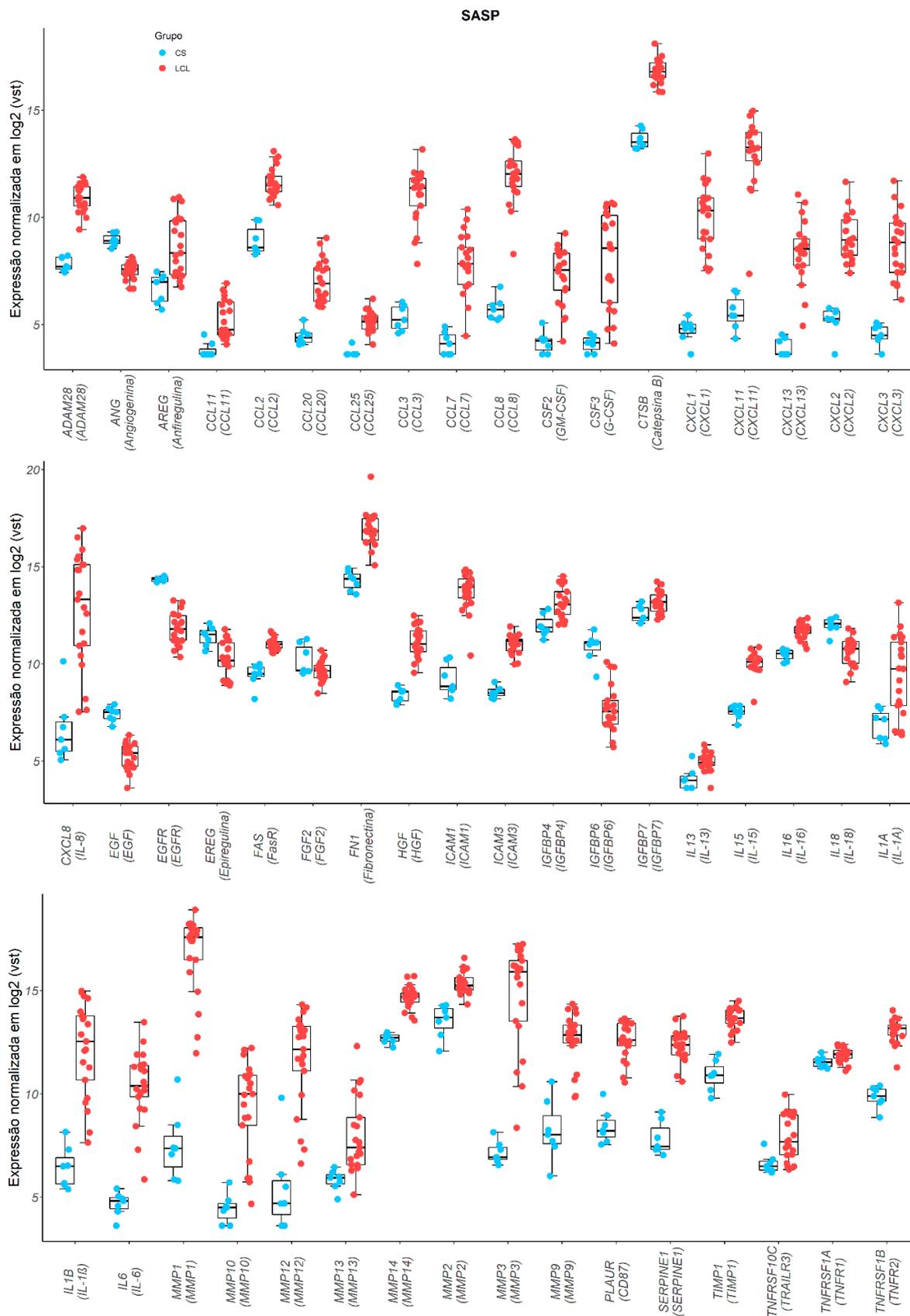


Figura 18 – Representação em diagrama de caixa com pontos dos genes pertencentes ao SASP comparando os grupos controle (CS, em azul, n = 7) com os pacientes (LCL, em vermelho, n = 21). Todos os genes desse gráfico são diferencialmente expressos, com valor de p ajustado de < 0,05.

5.5 DECONVOLUÇÃO DE CÉLULAS IMUNOLÓGICAS

O perfil de subtipos celulares presentes na infecção é um componente crucial para entender a resposta imunológica durante a infecção. Portanto, esse perfil foi avaliado durante a infecção por *L. braziliensis* através do transcriptoma total da lesão. Verificou-se um aumento em diversos subtipos de celulares nos pacientes em relação aos controles saudáveis (Apêndices C e D). Os principais subtipos aumentados na lesão foram os neutrófilos, monócitos, linfócitos T CD8⁺ EMRA, linfócitos T CD8 de memória efetora, células dendríticas e células NK CD16⁺ (Figura 19). Isto corrobora nossos achados anteriores que demonstram o acúmulo de células citotóxicas (CD8 e NK) nas lesões de pacientes com LC (COVRE et al., 2020).

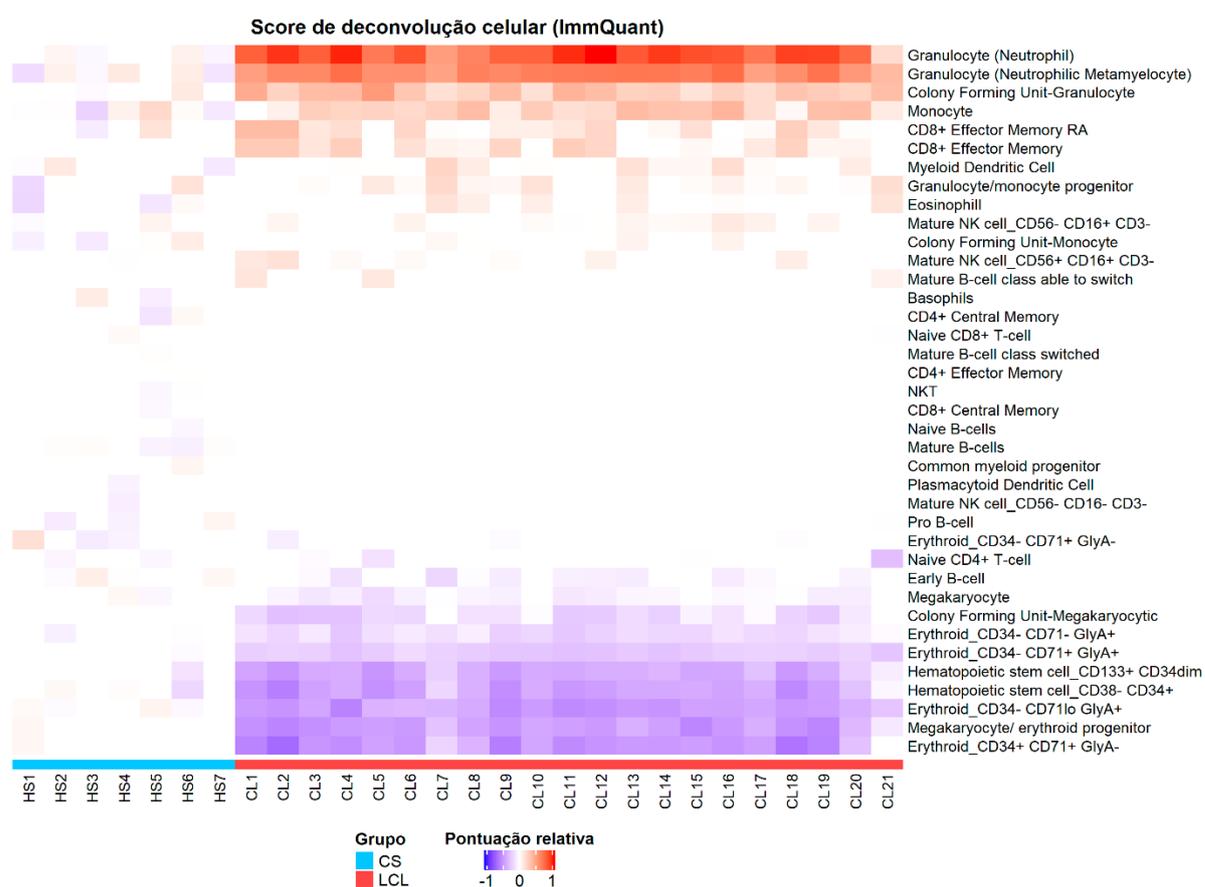


Figura 19 – Mapa de calor das pontuações relativas das populações células. Os tipos celulares foram identificados pelo programa ImmQuant para deconvolução de tipos celulares do sistema imunológico, comparando o grupo controle (CS, n = 7) e o grupo de pacientes (LCL, n = 21).

A partir do perfil celular encontrado, selecionamos as células altamente diferenciadas, características da senescência, para avaliar a sua relação com alguns dos genes analisados e expressos na lesão. Para isso, realizamos uma correlação de Spearman entre a expressão desses genes e as populações celulares selecionados. Os resultados demonstram que a presença de células T CD8⁺ de memória efetora, T CD8⁺ EMRA e células NK altamente diferenciadas (CD56⁺CD16⁺) foram positivamente correlacionadas com a maior parte dos genes associados as assinaturas transcricionais de secreção (SASP) e sinalização intracelular associado a senescência (ATF, ATM, PD-L1, p16, Fas-L, Granzima A, Granzima B, Granzima H, Granzima K, Perforina, IFN- γ , TNF- α , IL-12R, IL-15, IL-15R, IL-18R, IL-1A, IL-1 β , IL-2R, KLRG1, Sestrin 2, DR5/TRAILR2, TRAILR3). Além disso, observamos também uma correlação positiva entre o acúmulo destas populações altamente diferenciadas e a atividade citotóxica, reforçando nossos achados e corroborando para o papel desta atividade mediada por células senescentes para a imunopatogênese da LCL (Figura 20).

6. DISCUSSÃO

A imunopatogênese da LCL é caracterizada pela presença de uma resposta celular inflamatória crônica, mediada por células T CD8⁺ e envolvendo principalmente a produção de TNF- α e de moléculas efetoras citotóxicas como as granzimas e perforinas (BACELLAR et al., 2002; CARVALHO et al., 2012; LOUZIR et al., 1998; RIBEIRO-DE-JESUS et al., 1998; SCOTT; NOVAIS, 2016). Em adição a isso, nosso grupo demonstrou recentemente que linfócitos T e células NK altamente diferenciados se acumulam, na circulação e lesão, durante a infecção por *L. braziliensis*, e estão associadas ao processo de imunopatogênese da LCL. (COVRE et al., 2019, 2020). As células T senescentes foram identificadas predominantemente no compartimento de células efetoras de memória que reexpressam o CD45RA (T_{EMRA}). Essas células apresentaram maior expressão de CD57, KLRG1, p38 MAPK, fosforilação da proteína H2AX, menores telômeros e menor expressão do componente catalítico da telomerase (hTERT), além de perda da expressão dos receptores costimulatórios CD27 e CD28 e menor capacidade proliferativa (COVRE et al., 2019).

Considerando o impacto das células senescentes na imunopatogênese da LCL, e com o objetivo de ampliar a caracterização dessas células na lesão, investigamos aqui o perfil de expressão de genes associados a assinatura transcricional da imunossenescência no ambiente lesional. Avaliamos o perfil de expressão diferencial de genes entre os grupos de controle saudáveis e de pacientes, onde foram encontrados 12.306 genes diferencialmente expressos. Esse resultado contrasta com a análise feita pelo grupo que obteve o transcriptoma dessas amostras (AMORIM et al., 2019), cuja análise demonstrou apenas 4.255 genes diferencialmente expressos. Essa diferença possivelmente foi relacionada a dois motivos: a metodologia utilizada e os limites definidos, onde Amorim e colaboradores utilizaram o método do pacote *limma* para análise diferencial, em contraste com o método *DESeq2* aqui utilizado. Além disso, os limiares do estudo anterior, foram definidos como valor de p ajustado menor que 0,01 e \log_2 *fold change* maior que 1 ou menor que -1 (no mínimo o dobro ou metade na alteração do gene), sendo então mais restritivo. Nesse estudo, utilizamos o limite convencional para o valor de p ajustado (p menor que 0,05) e não limitamos o valor de *fold change* com a finalidade de avaliar o maior número de alterações estatisticamente válidas possível dentro do escopo dos genes de interesse.

Em nosso estudo, o resultado da análise dos componentes principais distinguiu adequadamente os dois grupos amostrais, corroborando com resultados anteriores (AMORIM et al., 2019). No entanto, comparando ambos os estudos, fomos capazes

de determinar uma maior variância amostral (72% versus 63,7%), possivelmente decorrente dos diferentes métodos de normalização e estabilização de variância das amostras utilizadas. Também, identificamos 794 processos biológicos associados aos genes diferencialmente expressos nos pacientes em relação ao grupo controle. Os principais processos biológicos representados foram associados ao sistema imunológico. Deste último, a categoria “Ativação de células T” foi a com maior representação estatística (valor de p aproximadamente = $2,28 \times 10^{-20}$). Os demais processos foram majoritariamente associados à diferenciação e ativação de leucócitos, em particular de linfócitos T, além da regulação de respostas inflamatórias, processos os quais já foram experimentalmente associados à resposta imunológica à LCL (BACELLAR et al., 2002; CARVALHO et al., 2012; LOUZIR et al., 1998; RIBEIRO-DE-JESUS et al., 1998; SCOTT; NOVAIS, 2016).

Em relação aos processos associados à senescência, as categorias “Senescência Replicativa” e “Envelhecimento”, foram identificadas através do total de genes diferencialmente expressos e foram representadas com significância estatística, sendo os valores de p 0.00911 e 0.0278 respectivamente. Essa análise funcional ampla, indicou que genes associados a imunossenescência poderiam estar expressos no local da lesão corroborando achados anteriores de nosso grupo e fomentando nossa hipótese de que células imunossenescentes podem ser extensivas ao sítio lesional e participar da imunopatogênese da LCL por meio de uma alta resposta inflamatória e citotóxica (COVRE et al., 2019, 2020).

A aquisição de características de senescência está associada a patogênese de uma série de doenças crônicas (BRENCHLEY et al., 2003; FREUND et al., 2010; FÜLÖP et al., 2016; OUYANG et al., 2003; PAWELEC; GOUTTEFANGEAS, 2006; PEREIRA; AKBAR, 2016). A fim de ampliar a caracterização das células senescentes presentes na lesão, selecionamos 179 genes relacionados a características fenotípicas e funcionais mediadas por essas células. Desses genes selecionados, identificamos 143 diferencialmente expressos relacionados principalmente as características de: controle do ciclo celular, SASP e citotoxicidade, todos previamente associados a células senescentes e a imunopatogênese da LCL. A análise de agrupamento hierárquico utilizando apenas esses 143 genes também foi capaz de distinguir com eficácia os dois grupos amostrais utilizados no estudo. Além disso, dentre esses genes apenas 10 tiveram valor de p ajustado menor que 0,05 e maior que 0,01 (*FOXO1*, *FOXO3*, *KIR2DL2*, *EREG*, *FGF2*, *IGFBP7*, *TNFRSF1A*, *ATF2*,

ATF4, *PRKAG2*), indicando que a maioria dos genes envolvidos nesse fenótipo possuem alta significância estatística nessa análise. Em relação à alteração na expressão gênica, a maioria dos genes apresentou \log_2 *fold change* maior do que 1 ou menor que -1. Esse conjunto de dados indica uma presença robusta de uma assinatura transcricional associada a senescência celular, altamente distinguível entre indivíduos saudáveis e pacientes infectados com *L. braziliensis*.

Dentre as características citadas, o bloqueio do ciclo celular é uma marca associada ao processo de senescência, refletindo diretamente na diminuição da capacidade proliferativa das células. Esse controle do crescimento celular é estabelecido principalmente através da inibição de ciclinas intracelulares por mecanismos dependentes da ativação das proteínas p53-p21 e p16 (MUNK et al., 2018; RUFINI et al., 2013). Em nosso estudo, embora houvesse um aumento na expressão dos genes de ciclinas, também detectamos um aumento nesses inibidores característicos do processo senescência celular.

Observamos também alterações nos genes codificantes para fatores de transcrição, com a diminuição dos genes FOXO (*forkhead box transcription factors Class O*) e o aumento nos genes *ZEB2* e *TBX21*, características que são consistentes com um perfil de diferenciação terminal previamente descrito em células T CD8⁺, e relacionadas à senescência desse compartimento celular (DOMINGUEZ et al., 2015; MURTAZA et al., 2017; PEREIRA; AKBAR, 2016). De modo complementar, nossos resultados ainda demonstram o aumento da maior parte dos genes de sinalização associada a senescência, com destaque para os genes *ATM* e *SESN2*. O gene *ATM* é ativado em condições de estresse onde ocorre dano à fita dupla de DNA e, conseqüentemente, ativa proteínas envolvidas no reconhecimento e reparo do material genético. Esse processo leva à interrupção do ciclo celular, o reparo do DNA, ou à apoptose da célula (SHILOH; ZIV, 2013).

Nossos resultados ainda demonstram um aumento em sestrin 2 (*SESN2*). Sestrins são uma família de proteínas sensoras de estresse celular altamente conservadas, as quais são transcricionalmente reguladas por p53 e fatores de transcrição *forkhead box* (FOX) (BUDANOV; LEE; KARIN, 2010). Constitutivamente elevadas em linfócitos T CD4⁺ e CD8⁺ senescentes, essas moléculas, em particular sestrin 2, estão estreitamente relacionadas com a reprogramação de linfócitos T, regulando vias de sinalização associadas à senescência como redução da proliferação celular, dano ao DNA e atividade do TCR (LANNA et al., 2017). Uma

caracterização mais abrangente da expressão de genes em linfócitos CD8⁺ T_{EMRA} em comparação com linfócitos *naive* de nosso grupo, identificou que essas células apresentam uma redução na expressão de genes associados ao complexo do TCR e de moléculas costimulatórias (CD27 e CD28), concomitante ao aumento de genes de receptores (em especial os genes das famílias KIR, NKG e KLR) e genes de moléculas adaptadoras (EAT-2, DAP12, SYK e SAP) previamente associados à células NK (PEREIRA et al., 2020), adquirindo um fenótipo inato. Com esse fenótipo, principalmente dependente da expressão do receptor NKG2D associado a molécula DAP12, induzida pela expressão de sestrin 2, os linfócitos CD8⁺ T senescentes adquirirão a capacidade de ativação e reconhecimento de células infectadas ou tumorais independente da ativação do TCR (PEREIRA et al., 2020).

Além das características supracitadas, células senescentes apresentam grande capacidade de secretarem um grupo comum de moléculas, com destaque para citocinas pró-inflamatórias e quimiocinas. Essa característica é conhecida como SASP, e envolve a alteração na expressão de diversos genes (CAMPISI, 2013; COPPÉ et al., 2008, 2010), incluindo citocinas e quimiocinas inflamatórias e imunomoduladoras, como IL-1 β , IL-6, IL-7, IL-8, CCL20 e CCL8 (COPPÉ et al., 2008). Nossos resultados demonstraram que a maior parte dos genes diferencialmente expressos relacionados ao SASP tiveram sua expressão aumentada, junto a eles, também foram elevados os genes para receptores dessas citocinas. Esses dados indicam que essas citocinas podem exercer um efeito local crônico, servindo de estímulo para a manutenção do fenótipo pró-inflamatório e senescente nas células, como um feedback positivo (RODIER et al., 2009). Além disso, podem também alterar a estrutura e função de células estromais não acometidas pela infecção/doença (PARRINELLO et al., 2005), exacerbando o dano tecidual e a gravidade da lesão. Nos genes do SASP, mostramos também o aumento da expressão do gene para a citocina IL-15 (*IL15*), bem como o aumento do receptor dessa citocina, *IL15RA*. O aumento da expressão desses genes foi demonstrado recentemente na LCL cujo bloqueio através do tofacitinib, um inibidor de JAK1 e JAK3 que bloqueia a sinalização através do receptor dessa citocina, foi capaz de diminuir a resposta patogênica na lesão através da diminuição na expressão de granzima B (NOVAIS; NGUYEN; SCOTT, 2020). Isso caracteriza o IL-15 como uma importante citocina na imunopatogênese dessa doença, reforçando o papel do SASP na manutenção da lesão.

Além disso, dentre os 71 genes analisados para o SASP, apenas 19 não apresentaram diferenças entre os grupos, e apenas 6 tiveram sua expressão reduzida (*FGF2*, *EREG*, *IL18*, *ANG*, *EGFR*, *EGF* e *IGFBP6*). Com exceção de *IL18* e *IGFBP6*, esses genes que foram regulados negativamente estão relacionados à mitogênese, angiogênese e a cicatrização de feridas (COHEN, 1987; HARDWICKE et al., 2008; LYONS et al., 2017; NUGENT; IOZZO, 2000; RIESE; CULLUM, 2014) e a diminuição na expressão desses genes pode, portanto, estar relacionada ao agravamento da lesão.

Ainda em suporte aos mecanismos associados à imunopatogênese, nossos achados demonstram uma alta expressão de genes relacionados à citotoxicidade mediada por granzima B e perforina e inflamação, que já foram descritos e previamente associados com a imunopatogênese da LCL (AMORIM et al., 2019; CHRISTENSEN et al., 2016; COVRE et al., 2019, 2020; FARIA et al., 2005, 2009; NOVAIS et al., 2015, 2017).

Além da grande atividade inflamatória e citotóxica, células T senescentes podem ser caracterizadas majoritariamente no compartimento de células de memória efetora que reexpressam CD45RA (T_{EMRA}) e são definidas pelo fenótipo $CD45RA^+CD45RO^-CD27^-CD28^-CCR7^-$ (BRENCHLEY et al., 2003; EFFROS, 1997; HENSON; RIDDELL; AKBAR, 2012; PEREIRA; AKBAR, 2016; TARAZONA et al., 2001; WENG; AKBAR; GORONZY, 2009; XU; LARBI, 2017). Uma análise por algoritmo da frequência relativa dos subtipos celulares presentes nas amostras de pele demonstrou que dentro dos subtipos celulares no ambiente lesional se encontram as células T $CD8^+$ Efetoras (*CD8+ Effector Memory*) e $T_{EMRA} CD8^+$ (*CD8+ Effector Memory RA*). A ocorrência dessas células também foi demonstrada anteriormente pelos autores do (AMORIM et al., 2019). No entanto, nosso estudo apresentou maior consistência nas frequências relativas dessas células, especialmente no grupo de pacientes. Em menor intensidade, também identificamos a presença de células NK altamente diferenciadas (*Mature NK cell_CD56+ CD16+ CD3-*). A identificação de uma população de linfócitos $T_{EMRA} CD8^+$ no ambiente lesional, embora de forma indireta, corrobora estudos anteriores de nosso grupo que demonstram o potencial migratório dessas células para a lesão (COVRE et al., 2019).

Por último, correlacionamos algumas populações celulares com a expressão de alguns genes importantes mediadores da senescência. As células $T_{EMRA} CD8^+$ foram positivamente e significativamente correlacionadas com a maioria dos genes testados para correlação, incluindo genes citotóxicos (Granzimas e Perforina), genes

inflamatórios (TNF- α e IFN- γ), genes receptores de citocinas, o receptor de diferenciação celular *KLRG1*, e genes associados à senescência como a proteína p16 (*CDKN2A*), ao *ATM* e ao Sestrin 2 (*SESN2*). As células T CD8⁺ Efetoras (*CD8+ Effector Memory*) NK altamente diferenciadas (*Mature NK cell_CD56+ CD16+ CD3-*) também se correlacionaram com esses genes, como p16 e ATM, porém em menor intensidade, quantidade e significância estatística. Esse resultado indica que as células lesionais apresentam um extensivo fenótipo senescente, representado principalmente por linfócitos T CD8⁺ e células NK terminalmente diferenciadas, que apresentam mecanismos citotóxicos e inflamatórios associados às lesões.

Coletivamente, os resultados do presente estudo demonstram uma forte participação de mecanismos associados à imunossenescência no ambiente lesional durante a LCL causada por *L. braziliensis*, corroborando resultados anteriores, e representando uma implicação direta sobre os eventos que caracterizam a imunopatogênese dessa doença.

Estudos do transcriptoma global são de extrema importância. Esses estudos são capazes de fornecer uma visão ampla dos mecanismos gênicos ativos no ambiente estudado. Aqui, associamos a expressão de genes relacionados à senescência no ambiente lesional da LCL. Entretanto, esse estudo possui algumas limitações. Embora o método escolhido (DESeq2) controle internamente para erros do tipo I (LOVE; HUBER; ANDERS, 2014), esse método é mais sensível a mudanças menores na expressão de genes (ZHOU; LINDSAY; ROBINSON, 2014), podendo levar ao aparecimento de falsos positivos ou à detecção de alterações muito baixas na expressão de genes como significantes estatisticamente. Isso, em conjunto com os limites mais permissivos que utilizamos pode explicar a grande quantidade de genes diferencialmente expressos encontrados, além da baixa alteração encontrada em alguns genes, como por exemplo genes de fatores de transcrição (genes *FOXO*), genes de sinalização (genes *ATF* e genes de p-AMPK) e genes do SASP (genes *EREG*, *FGF2*, *IGFBP7* e *TNFRSF1A*).

Outra limitação desse estudo é que essa análise de transcriptoma global representa a expressão gênica de um tecido heterogêneo, composto por diversos tipos celulares. Assim, embora sejamos capazes de prever os subtipos celulares presentes no local da lesão e sua associação com a expressão de determinados genes, essa se constitui como uma inferência indireta e que não é capaz de discriminar com precisão absoluta os mecanismos de senescência e quais subtipos

celulares são responsáveis por esses mecanismos. Entretanto, a maior parte dos genes analisados sofreu grandes diferenças na sua expressão (maior que o dobro, para genes regulados positivamente, ou menor que a metade, para genes regulados negativamente), o que indica uma forte presença dos mecanismos associados à imunossenescência aqui estudados, mesmo em um estudo de população celular heterogênea. Estudos envolvendo a análise do transcriptoma de células senescentes, através do sequenciamento de RNA de células únicas (*single-cell RNA-Seq*), isoladas diretamente do ambiente lesional, constituem uma perspectiva futura de investigar com maior resolução os fatores possivelmente associados à essa imunopatogênese.

7. CONCLUSÃO

Os resultados encontrados nesse trabalho permitem concluir sobre a expressão gênica na lesão durante a LCL causada por *L. braziliensis*:

1. A análise global de expressão diferencial revelou 12.306 genes diferencialmente expressos, dos quais 6.076 foram regulados positivamente e 6.230 foram regulados negativamente;
2. Nos genes diferencialmente expressos, encontramos um perfil transcricional de característica inflamatória, citotóxica e senescente. Esse perfil senescente é dominado por genes que controlam a diferenciação e proliferação celular e principalmente por genes associados ao fenótipo secretor associado a senescência - SASP;
3. A análise funcional demonstrou a representação de processos biológicos majoritariamente associados à diferenciação e ativação de linfócitos, resposta inflamatória e secreção de citocinas, sendo também associados processos relativos ao envelhecimento e à senescência replicativa;
4. O perfil celular relativo encontrado na lesão, associa as características de senescência à Linfócitos T_{EMRA} CD8⁺ e células NK altamente diferenciadas, as quais foram positivamente correlacionadas com genes de citotoxicidade, inflamação e senescência, indicando a capacidade dessas células em mediar a imunopatogênese da LCL.

REFERÊNCIAS

ABDELLAH, Z. et al. Finishing the euchromatic sequence of the human genome. **Nature**, v. 431, n. 7011, p. 931–945, 2004.

AKBAR, A. N.; FLETCHER, J. M. Memory T cell homeostasis and senescence during aging. **Current Opinion in Immunology**, v. 17, n. 5, p. 480–485, 2005.

AKBAR, A. N.; HENSON, S. M. Are senescence and exhaustion intertwined or unrelated processes that compromise immunity? **Nature Reviews Immunology**, v. 11, n. 4, p. 289–295, 2011. Disponível em: <<http://dx.doi.org/10.1038/nri2959>>.

AKBAR, A. N.; HENSON, S. M.; LANNA, A. Senescence of T Lymphocytes: Implications for Enhancing Human Immunity. **Trends in Immunology**, v. 37, n. 12, p. 866–876, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.it.2016.09.002>>.

ALVAR, J. et al. Leishmaniasis worldwide and global estimates of its incidence. **PLoS ONE**, v. 7, n. 5, 2012.

AMORIM, C. F. et al. Variable gene expression and parasite load predict treatment outcome in cutaneous leishmaniasis. **Science Translational Medicine**, v. 11, n. 519, p. 1–10, 2019.

ANDERSON, S. Shotgun DNA sequencing using cloned DNase I-generated fragments. **Nucleic Acids Research**, v. 9, n. 13, p. 3015–3027, 1981.

ANTONELLI, L. R. V. et al. Activated inflammatory T cells correlate with lesion size in human cutaneous leishmaniasis. **Immunology Letters**, v. 101, n. 2, p. 226–230, 2005.

BACELLAR, O. et al. Up-Regulation of Th1-Type Responses in Mucosal Leishmaniasis Patients. **Infection and Immunity**, v. 70, n. 12, p. 6734–6740, 1 dez. 2002. Disponível em: <<http://iai.asm.org/cgi/doi/10.1128/IAI.70.12.6734-6740.2002>>.

BALABAN, R. S.; NEMOTO, S.; FINKEL, T. Mitochondria, oxidants, and aging. **Cell**, v. 120, n. 4, p. 483–495, 2005.

BARRETT, T. et al. NCBI GEO: Archive for functional genomics data sets - Update. **Nucleic Acids Research**, v. 41, n. D1, p. 991–995, 2013.

BELKAID, Y. et al. CD8 + T Cells Are Required for Primary Immunity in C57BL/6 Mice Following Low-Dose, Intradermal Challenge with *Leishmania major*. **The Journal of Immunology**, v. 168, n. 8, p. 3992–4000, 2002.

BEN-PORATH, I.; WEINBERG, R. A. The signals and pathways activating cellular senescence. **International Journal of Biochemistry and Cell Biology**, v. 37, n. 5 SPEC. ISS., p. 961–976, 2005.

_____. When cells get stressed: An integrative view of cellular senescence. **Journal of Clinical Investigation**, v. 113, n. 1, p. 8–13, 2004.

BLACKBURN, E. H. Telomeres and telomerase: Their mechanisms of action and the effects of altering their functions. **FEBS Letters**, v. 579, n. 4 SPEC. ISS., p. 859–862, 2005.

BOYLE, E. I. et al. GO::TermFinder - Open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. **Bioinformatics**, v. 20, n. 18, p. 3710–3715, 2004.

BRAIG, M.; SCHMITT, C. A. Oncogene-induced senescence: Putting the brakes on tumor development. **Cancer Research**, v. 66, n. 6, p. 2881–2884, 2006.

BRENCHLEY, J. M. et al. Expression of CD57 defines replicative senescence and antigen-induced apoptotic death of CD8⁺ T cells. **Blood**, v. 101, n. 7, p. 2711–2720, 2003.

BUDANOV, A. V.; LEE, J. H.; KARIN, M. Stressin' Sestrins take an aging fight. **EMBO Molecular Medicine**, v. 2, n. 10, p. 388–400, 2010.

CAMPISI, J. Aging, Cellular Senescence, and Cancer. **Annual Review of Physiology**, v. 75, n. 1, p. 685–705, 2013.

_____. Cellular senescence: A link between cancer and age-related degenerative disease? **Seminars in Cancer Biology**, v. 23, n. 1, p. 1–7, set. 2011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1044579X11000502>>.

_____. Senescent cells, tumor suppression, and organismal aging: Good citizens, bad neighbors. **Cell**, v. 120, n. 4, p. 513–522, 2005.

CAMPISI, J.; D'ADDA DI FAGAGNA, F. Cellular senescence: When bad things happen to good cells. **Nature Reviews Molecular Cell Biology**, v. 8, n. 9, p. 729–740, 2007.

CARDOSO, T. M. et al. Protective and pathological functions of CD8⁺ T cells in *Leishmania braziliensis* infection. **Infection and Immunity**, v. 83, n. 3, p. 898–906, 2015.

CARVALHO, E. M. et al. Clinical and immunopathological aspects of disseminated cutaneous leishmaniasis. **Acta Tropica**, v. 56, n. 4, p. 315–325, 1994.

CARVALHO, L. P. et al. Protective and pathologic immune responses in human tegumentary leishmaniasis. **Frontiers in Immunology**, v. 3, n. OCT, p. 1–9, 2012.

CASTELLANO, L. R. et al. Th1/Th2 immune responses are associated with active cutaneous leishmaniasis and clinical cure is associated with strong interferon- γ production. **Human Immunology**, v. 70, n. 6, p. 383–390, 2009. Disponível em: <<http://dx.doi.org/10.1016/j.humimm.2009.01.007>>.

CHIEN, K. R.; KARSENTY, G. Longevity and lineages: Toward the integrative biology of degenerative diseases in heart, muscle, and bone. **Cell**, v. 120, n. 4, p. 533–544, 2005.

CHRISTENSEN, S. M. et al. Meta-transcriptome Profiling of the Human-*Leishmania braziliensis* Cutaneous Lesion. **PLoS Neglected Tropical Diseases**, v. 10, n. 9, p. 1–17, 2016.

COHEN, S. Epidermal growth factor. **In Vitro Cellular & Developmental Biology**, v. 23, n. 4, p. 239–246, abr. 1987. Disponível em: <<http://link.springer.com/10.1007/BF02623704>>.

COPPÉ, J.-P. et al. Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. **PLoS biology**, v. 6, n. 12, 2008.

_____. The Senescence-Associated Secretory Phenotype: The Dark Side of Tumor Suppression. **Annual Review of Pathology: Mechanisms of Disease**, v. 5, n. 1, p. 99–118, 2010.

COSTA, J. M. L. et al. Cura espontânea da leishmaniose causada por *Leishmania Viannia Braziliensis* em lesões cutâneas. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 23, n. 4, p. 205–208, dez. 1990. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0037-86821990000400004&lng=pt&tlng=pt>.

COVRE, L. P. et al. Circulating senescent T cells are linked to systemic inflammation and lesion size during human cutaneous leishmaniasis. **Frontiers in Immunology**, v. 10, n. JAN, p. 1–12, 2019.

_____. Compartmentalized cytotoxic immune response leads to distinct pathogenic roles of natural killer and senescent CD8+ T cells in human cutaneous leishmaniasis. **Immunology**, p. 1–12, 2020.

CUNHA, C. F. et al. Contribution of *Leishmania braziliensis* antigen-specific CD4+ T, CD8+ T, NK and CD3+CD56+NKT cells in the immunopathogenesis of cutaneous leishmaniasis patients: Cytotoxic, activation and exhaustion profiles. **PLoS ONE**, v. 15, n. 3, p. 1–15, 2020.

D'ADDA DI FAGAGNA, F. Living on a break: Cellular senescence as a DNA-damage response. **Nature Reviews Cancer**, v. 8, n. 7, p. 512–522, 2008.

DA-CRUZ, A. M. et al. Flow cytometric analysis of cellular infiltrate from American tegumentary leishmaniasis lesions. **British Journal of Dermatology**, v. 153, n. 3, p. 537–543, 2005.

_____. *Leishmania*-reactive CD4+ and CD8+ T cells associated with cure of human cutaneous leishmaniasis. **Infection and Immunity**, v. 62, n. 6, p. 2614–2618, 1994.

DA SILVA SANTOS, C.; BRODSKY, C. I. The role of CD4 and CD8 T cells in human cutaneous leishmaniasis. **Frontiers in Public Health**, v. 2, n. SEP, p. 1–6, 2014.

DI LEONARDO, A. et al. DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts. **Genes and Development**, v. 8, n. 21, p. 2540–2551, 1994.

DOMINGUEZ, C. X. et al. The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. **Journal of Experimental Medicine**, v. 212, n. 12, p. 2041–2056, 2015.

DONG, D. et al. Structural basis of assembly of the human T cell receptor–CD3 complex. **Nature**, v. 573, n. 7775, p. 546–552, 2019. Disponível em: <<http://dx.doi.org/10.1038/s41586-019-1537-0>>.

EFFROS, R. B. Loss of CD28 expression on T lymphocytes: A marker of replicative senescence. **Developmental and Comparative Immunology**, v. 21, n. 6, p. 471–478, 1997.

FAGNONI, F. F. et al. Shortage of circulating naive CD8+ T cells provides new insights on immunodeficiency in aging. **Blood**, v. 95, n. 9, p. 2860–2868, 2000.

FARIA, D. R. et al. Decreased in situ expression of interleukin-10 receptor is correlated with the exacerbated inflammatory and cytotoxic responses observed in mucosal leishmaniasis. **Infection and Immunity**, v. 73, n. 12, p. 7853–7859, 2005.

_____. Recruitment of CD8+ T cells expressing granzyme A is associated with lesion progression in human cutaneous leishmaniasis. **Parasite Immunology**, v. 31, n. 8, p. 432–439, 2009.

FRANCESCHI, C.; MONTI, D.; BARBIERI, D.; et al. Immunosenescence in Humans: Deterioration or Remodelling? **International Reviews of Immunology**, v. 12, n. 1, p. 57–74, 10 jan. 1995. Disponível em: <<http://www.tandfonline.com/doi/full/10.3109/08830189509056702>>.

FRANCESCHI, C. et al. Inflamm-aging: An Evolutionary Perspective on Immunosenescence. **Annals of the New York Academy of Sciences**, v. 908, n. 1, p. 244–254, 25 jan. 2006. Disponível em: <<http://doi.wiley.com/10.1111/j.1749-6632.2000.tb06651.x>>.

FRANCESCHI, C.; MONTI, D.; SANSONI, P.; et al. The immunology of exceptional individuals: the lesson of centenarians. **Immunology Today**, v. 16, n. 1, p. 12–16, jan. 1995. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0167569995800646>>.

FRANCESCHI, C.; BONAFÈ, M.; VALENSIN, S. Human immunosenescence: the prevailing of innate immunity, the failing of clonotypic immunity, and the filling of immunological space. **Vaccine**, v. 18, n. 16, p. 1717–1720, 10 fev. 2000. Disponível em: <<http://www.tandfonline.com/doi/full/10.3109/08830189509056702>>.

FREUND, A. et al. Inflammatory networks during cellular senescence: causes and consequences. **Trends in Molecular Medicine**, v. 16, n. 5, p. 238–246, 2010.

FREUND, A.; PATIL, C. K.; CAMPISI, J. P38MAPK is a novel DNA damage response-independent regulator of the senescence-associated secretory phenotype. **EMBO Journal**, v. 30, n. 8, p. 1536–1548, 2011. Disponível em: <<http://dx.doi.org/10.1038/emboj.2011.69>>.

FRISHBERG, A. et al. ImmQuant: a user-friendly tool for inferring immune cell-type composition from gene-expression data. **Bioinformatics**, v. 32, n. 24, p. 3842–3843, 15 dez. 2016. Disponível em: <<https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw535>>.

FÜLÖP, T. et al. The role of immunosenescence in the development of age-related diseases. **Revista de Investigacion Clinica**, v. 68, n. 2, p. 84–91, 2016.

GALILI, T. dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. **Bioinformatics**, v. 31, n. 22, p. 3718–3720, 2015.

GINALDI, L. et al. Immunosenescence and infectious diseases. **Microbes and Infection**, v. 3, n. 10, p. 851–857, 2001.

GU, Z.; EILS, R.; SCHLESNER, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. **Bioinformatics**, v. 32, n. 18, p. 2847–2849, 2016.

HARDWICKE, J. et al. Epidermal growth factor therapy and wound healing - Past, present and future perspectives. **Surgeon**, v. 6, n. 3, p. 172–177, 2008. Disponível em: <[http://dx.doi.org/10.1016/S1479-666X\(08\)80114-X](http://dx.doi.org/10.1016/S1479-666X(08)80114-X)>.

HAYFLICK, L. The limited in vitro lifetime of human diploid cell strains. **Experimental Cell Research**, v. 37, n. 3, p. 614–636, 1965.

HAYFLICK, L.; MOORHEAD, P. S. The serial cultivation of human diploid cell strains. **Experimental Cell Research**, v. 25, n. 3, p. 585–621, dez. 1961. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/0014482761901926>>.

HENSON, S. M.; RIDDELL, N. E.; AKBAR, A. N. Properties of end-stage human T cells defined by CD45RA re-expression. **Current Opinion in Immunology**, v. 24, n. 4, p. 476–481, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.coi.2012.04.001>>.

IWASA, H.; HAN, J.; ISHIKAWA, F. Mitogen-activated protein kinase p38 defines the common senescence-signalling pathway. **Genes to Cells**, v. 8, n. 2, p. 131–144, fev. 2003. Disponível em: <<http://doi.wiley.com/10.1046/j.1365-2443.2003.00620.x>>.

KAYE, P.; SCOTT, P. Leishmaniasis: Complexity at the host-pathogen interface. **Nature Reviews Microbiology**, v. 9, n. 8, p. 604–615, 2011. Disponível em: <<http://dx.doi.org/10.1038/nrmicro2608>>.

LANDER, E. S. et al. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860–921, 2001.

LANNA, A. et al. A sestrin-dependent Erk–Jnk–p38 MAPK activation complex inhibits immunity during aging. **Nature Immunology**, v. 18, n. 3, p. 354–363, 23 mar. 2017. Disponível em: <<http://www.nature.com/articles/ni.3665>>.

_____. The kinase p38 activated by the metabolic regulator AMPK and scaffold TAB1 drives the senescence of human T cells. **Nature Immunology**, v. 15, n. 10, p. 965–972, 2014. Disponível em: <<http://dx.doi.org/10.1038/ni.2981>>.

LINTON, P. J.; DORSHKIND, K. Age-related changes in lymphocyte development and function. **Nature Immunology**, v. 5, n. 2, p. 133–139, 2004.

LOMBARD, D. B. et al. DNA repair, genome stability, and aging. **Cell**, v. 120, n. 4, p. 497–512, 2005.

LOUZIR, H. et al. Immunologic determinants of disease evolution in localized cutaneous leishmaniasis due to *Leishmania major*. **Jid**, v. 177, p. 1687–1695, 1998.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 1–21, 2014.

LOWE, R. et al. Transcriptomics technologies. **PLoS Computational Biology**, v. 13, n. 5, p. 1–23, 2017.

LYNCH, H. E. et al. Thymic involution and immune reconstitution. **Trends in Immunology**, v. 30, n. 7, p. 366–373, 2009.

LYONS, S. M. et al. RNA biology of angiogenin: Current state and perspectives. **RNA Biology**, v. 14, n. 2, p. 171–178, 2017. Disponível em: <<http://dx.doi.org/10.1080/15476286.2016.1272746>>.

MARETTI-MIRA, A. C. et al. Transcriptome Patterns from Primary Cutaneous Leishmania braziliensis Infections Associate with Eventual Development of Mucosal Disease in Humans. **PLoS Neglected Tropical Diseases**, v. 6, n. 9, 2012.

MARTIN, G. M. Genetic modulation of senescent phenotypes in Homo sapiens. **Cell**, v. 120, n. 4, p. 523–532, 2005.

MELBY, P. C. et al. Increased expression of proinflammatory cytokines in chronic lesions of human cutaneous leishmaniasis. **Infection and Immunity**, v. 62, n. 3, p. 837–842, 1994.

METZKER, M. L. Sequencing technologies: the next generation. **Nature Reviews Genetics**, v. 11, n. 1, p. 31–46, 2010. Disponível em: <<http://dx.doi.org/10.1038/nrg2626>>.

MILLER, M. B.; TANG, Y. W. Basic concepts of microarrays and potential applications in clinical microbiology. **Clinical Microbiology Reviews**, v. 22, n. 4, p. 611–633, 2009.

MINISTÉRIO DA SAÚDE. **Manual De Vigilância Da Leishmaniose Tegumentar**. [S.l.: s.n.], 2017. Disponível em: <<http://pesquisa.bvsalud.org/bvsm/resouce/pt/mis-38798#.WmSchsQMGLw.mendeley>>.

MUNK, R. et al. **Senescence-Associated MicroRNAs**. 1. ed. [S.l.]: Elsevier Inc., 2018. v. 334. Disponível em: <<http://dx.doi.org/10.1016/bs.ircmb.2017.03.008>>.

MUÑOZ-ESPÍN, D.; SERRANO, M. Cellular senescence: From physiology to pathology. **Nature Reviews Molecular Cell Biology**, v. 15, n. 7, p. 482–496, 2014. Disponível em: <<http://dx.doi.org/10.1038/nrm3823>>.

MURPHY, K.; WEAVER, C. **Janeway's Immunobiology**. 9th. ed. New York, NY: Garland Science/Taylor & Francis, 2017.

MURTAZA, G. et al. FOXO Transcriptional Factors and Long-Term Living. **Oxidative Medicine and Cellular Longevity**, v. 2017, 2017.

NOVAIS, F. O. et al. CD8+T cell cytotoxicity mediates pathology in the skin by inflammasome activation and IL-1 β production. **PLoS Pathogens**, v. 13, n. 2, p. 1–21, 2017.

_____. CD8 + T Cells Lack Local Signals To Produce IFN- γ in the Skin during Leishmania Infection. **The Journal of Immunology**, p. ji1701597, 24 jan. 2018. Disponível em: <<http://www.jimmunol.org/lookup/doi/10.4049/jimmunol.1701597>>.

_____. Genomic Profiling of Human *Leishmania braziliensis* Lesions Identifies Transcriptional Modules Associated with Cutaneous Immunopathology. **Journal of Investigative Dermatology**, v. 135, n. 1, p. 94–101, jan. 2015. Disponível em: <file:///C:/Users/Carla Carolina/Desktop/Artigos para acrescentar na qualificação/The impact of birth weight on cardiovascular disease risk in the.pdf>.

NOVAIS, F. O.; NGUYEN, B. T.; SCOTT, P. Granzyme B inhibition by tofacitinib blocks pathology induced by CD8 T cells in cutaneous leishmaniasis. **The Journal of Investigative Dermatology**, 2020. Disponível em: <<https://doi.org/10.1016/j.jid.2020.07.011>>.

NOVAIS, F. O.; SCOTT, P. CD8+ T cells in cutaneous leishmaniasis: the good, the bad, and the ugly. **Seminars in Immunopathology**, v. 37, n. 3, p. 251–259, 2015.

NOVERSHTERN, N. et al. Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis. **Cell**, v. 144, n. 2, p. 296–309, jan. 2011. Disponível em: <<http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=21241896&retmode=ref&cmd=prlinks%0Apapers3://publication/doi/10.1016/j.cell.2011.01.004>>.

NUGENT, M. A.; IOZZO, R. V. Fibroblast growth factor-2. **International Journal of Biochemistry and Cell Biology**, v. 32, n. 2, p. 115–120, 2000.

OPAS. Leishmaniose cutânea e mucosa 2018. **Informe Leishmanioses**, p. 1, 2019. Disponível em: <www.paho.org/leishmaniasis>.

OSHLACK, A.; ROBINSON, M. D.; YOUNG, M. D. From RNA-seq reads to differential expression results. **Genome Biology**, v. 11, n. 12, p. 1–10, 2010.

OUYANG, Q. et al. Age-associated accumulation of CMV-specific CD8+ T cells expressing the inhibitory killer cell lectin-like receptor G1 (KLRG1). **Experimental Gerontology**, v. 38, n. 8, p. 911–920, 2003.

PARRINELLO, S. et al. Stromal-epithelial interactions in aging and cancer: Senescent fibroblasts alter epithelial cell differentiation. **Journal of Cell Science**, v. 118, n. 3, p. 485–496, 2005.

PAWELEC, G.; GOUTTEFANGEAS, C. T-cell dysregulation caused by chronic antigenic stress: The role of CMV in immunosenescence? **Aging Clinical and Experimental Research**, v. 18, n. 2, p. 171–173, 2006.

PEREIRA, B. I. et al. Sestrins induce natural killer function in senescent-like CD8+ T cells. **Nature Immunology**, v. 21, n. 6, p. 684–694, 2020. Disponível em: <<http://dx.doi.org/10.1038/s41590-020-0643-3>>.

PEREIRA, B. I.; AKBAR, A. N. Convergence of innate and adaptive immunity during human aging. **Frontiers in Immunology**, v. 7, n. NOV, p. 1–9, 2016.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. **R Foundation for Statistical Computing**. Vienna, Austria: [s.n.]. Disponível em: <<https://www.r-project.org/>>. , 2020

RIBEIRO-DE-JESUS, A. et al. Cytokine profile and pathology in human leishmaniasis. **Brazilian Journal of Medical and Biological Research**, v. 31, n. 1, p. 143–148, 1998.

RIESE, D. J.; CULLUM, R. L. Epiregulin: Roles in normal physiology and cancer. **Seminars in Cell & Developmental Biology**, v. 28, n. 1, p. 49–56, abr. 2014. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf>>.

RODIER, F. et al. Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. **Nature Cell Biology**, v. 11, n. 8, p. 973–979, 13 ago. 2009. Disponível em: <<http://www.nature.com/articles/ncb1909>>.

RSTUDIO TEAM. **Integrated Development for R. RStudio**. Disponível em: <<https://www.rstudio.com/>>. Acesso em: 12 jun. 2018.

RUFINI, A. et al. Senescence and aging: The critical roles of p53. **Oncogene**, v. 32, n. 43, p. 5129–5143, 2013.

SALMAN, S. M.; RUBEIZ, N. G.; KIBBI, A. G. Cutaneous leishmaniasis: Clinical features and diagnosis. **Clinics in Dermatology**, v. 17, n. 3, p. 291–296, 1999.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors. **Proceedings of the National Academy of Sciences**, v. 74, n. 12, p. 5463–5467, 1 dez. 1977. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.74.12.5463>>.

SANTOS, C. D. S. et al. CD8+ granzyme B+ -mediated tissue injury vs. CD4 + IFN γ + -mediated parasite killing in human cutaneous leishmaniasis. **Journal of Investigative Dermatology**, v. 133, n. 6, p. 1533–1540, 2013.

SCHUSTER, S. C. Next-generation sequencing transforms today's biology. **Nature Methods**, v. 5, n. 1, p. 16–18, 2008.

SCORZA, B. M.; CARVALHO, E. M.; WILSON, M. E. Cutaneous manifestations of human and murine leishmaniasis. **International Journal of Molecular Sciences**, v. 18, n. 6, 2017.

SCOTT, P.; NOVAIS, F. O. Cutaneous leishmaniasis: Immune responses in protection and pathogenesis. **Nature Reviews Immunology**, v. 16, n. 9, p. 581–592, 2016.

SEAN, D.; MELTZER, P. S. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. **Bioinformatics**, v. 23, n. 14, p. 1846–1847, 2007.

SHILOH, Y.; ZIV, Y. The ATM protein kinase: Regulating the cellular response to genotoxic stress, and more. **Nature Reviews Molecular Cell Biology**, v. 14, n. 4, p. 197–210, 2013. Disponível em: <<http://dx.doi.org/10.1038/nrm3546>>.

SINAN. Casos de Leishmaniose Tegumentar Americana. Brasil, Grandes Regiões e Unidades Federadas. 1990 a 2018. **Ministério da Saúde**, p. 2008–2008, 2019.

STADEN, R. A strategy of DNA sequencing employing computer programs. **Nucleic Acids Research**, v. 6, n. 7, p. 2601–2610, 1979. Disponível em: <<https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/6.7.2601>>.

TARAZONA, R. et al. Increased expression of NK cell markers on T lymphocytes in aging and chronic activation of the immune system reflects the accumulation of effector/senescent T cells. **Mechanisms of Ageing and Development**, v. 121, n. 1–3, p. 77–88, 2001.

UNGER, A. et al. Association of treatment of American cutaneous leishmaniasis prior to ulcer development with high rate of failure in northeastern Brazil. **The American journal of tropical medicine and hygiene**, v. 80, n. 4, p. 574–9, abr. 2009. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf>>.

UZONNA, J. E.; JOYCE, K. L.; SCOTT, P. Low dose Leishmania major promotes a transient T helper cell type 2 response that is down-regulated by interferon γ -producing CD8+ T cells. **Journal of Experimental Medicine**, v. 199, n. 11, p. 1559–1566, 2004.

VENTER, J. C. et al. The Sequence of the Human Genome. **Science**, v. 291, n. 5507, p. 1304–1351, 16 fev. 2001. Disponível em: <<https://www.sciencemag.org/lookup/doi/10.1126/science.1058040>>.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, n. 1, p. 57–63, 1 jan. 2009. Disponível em: <<http://www.nature.com/articles/nrg2484>>.

WENG, N.; AKBAR, A. N.; GORONZY, J. CD28– T cells: their role in the age-associated decline of immune function. **Trends in Immunology**, v. 30, n. 7, p. 306–312, jul. 2009. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1471490609000982>>.

WHO. Leishmaniasis - Epidemiological Report of the Americas. p. 2–5, 2019.

_____. Leishmaniasis in high-burden countries: an epidemiological update based on data reported in 2014. **Weekly epidemiological record**, v. 91, n. 22, p. 287–296, 2016.

_____. Unveiling the neglect of leishmaniasis. p. 2017, 2017.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. New York, NY: Springer-Verlag New York, 2016.

_____. Welcome to the Tidyverse. **Journal of Open Source Software**, v. 4, n. 43, p. 1686, 21 nov. 2019. Disponível em: <<https://joss.theoj.org/papers/10.21105/joss.01686>>. Acesso em: 21 abr. 2020.

WICKHAM, H.; HESTER, J.; FRANCOIS, R. Read Rectangular Text Data [R package readr version 1.3.1]. 2018. Disponível em: <<https://cran.r-project.org/package=readr>>. Acesso em: 21 abr. 2020.

XU, W.; LARBI, A. Markers of T cell senescence in humans. **International Journal of Molecular Sciences**, v. 18, n. 8, p. 1–13, 2017.

YU, G. et al. clusterProfiler: An R package for comparing biological themes among gene clusters. **OMICS A Journal of Integrative Biology**, v. 16, n. 5, p. 284–287, 2012.

_____. **enrichplot: Visualization of Functional Enrichment Result [R package version 1.4.0]**. Disponível em: <<https://github.com/GuangchuangYu/enrichplot>>. Acesso em: 21 abr. 2020.

ZGLINICKI, T. Von et al. Human cell senescence as a DNA damage response. **Mechanisms of Ageing and Development**, v. 126, n. 1, p. 111–117, 2005.

ZHOU, X.; LINDSAY, H.; ROBINSON, M. D. Robustly detecting differential expression in RNA sequencing data using observation weights. **Nucleic Acids Research**, v. 42, n. 11, 2014.

APÊNDICES

APÊNDICE A – CATEGORIAS ADICIONAIS SOBRRERREPRESENTADAS NA ONTOLOGIA GÊNICA

Termo GO	Descrição	Nº de Genes	Valor de p	p ajustado	Valor de q	Genes envolvidos
GO:0090399	Senescência Replicativa	13	7.11E-04	9.11E-03	7.26E-03	<i>SERPINE1, MME, CDKN2A, TERT, CDKN1A, ATM, CHEK1, CTC1, CHEK2, ATR, TP53, WNT16, PLA2R1</i> <i>KIR2DL4, BCL2A1, ICAM1, SERPINE1, PAX5, KMO, KYNU, SOD2, PTGS2, NTRK1, ITGB2, SLC32A1, HLA-G, C1QA, CD68, MPO, IL10, WNT1, CTSC, TNFRSF1B, UCP2, VCAM1, NAMPT, IL15, TIMP1, MME, PITX3, KRT16, GJB2, LRRK2, INPP5D, VASH1, TGFB1, NFKB2, CDKN2A, ERO1A, ENG, PML, SERPING1, GRB2, LITAF, ADM, TERT, ADA, HMGA2, LOXL2, TYMS, TGFB3, HAMP, NEK6, BAK1, CASP9, COL4A2, AURKB, BRCA2, BGLAP, HTR2A, GJB6, CDKN1A, ENO3, CALR, H2AFX, EDN1, FBXO5, LIMS1, AGER, KCNMB1, TBX2, FOXM1, ATM, CX3CL1, NUP62, RAD54L, NOX4, CDK1, JUN, PDE4D, RELA, STAT3, APAF1, RPN2, MAP3K3, ID2, IRAK1, RBL1, HYAL2, CISD2, ZMIZ1, CHEK1, PRKCD, CTC1, CHEK2, SERP1, ATG7, UCP3, MNT, CNP, SRF, CARM1, FZR1, PNPT1, AKT1, TGFB2, GCLM, ATR, PICALM, JUND, ILK, HTRA2, PTEN, POLG, NEK4, MORC3, MTOR, SCAP, DNAJA3, ZNF277, SIRT3, TP53, SIRT1, NDUFS6, RWDD1, WRN, BCL2, CTNNA1, SOD1, NFE2L2, ZKSCAN3, RSL1D1, EDNRB, ATP8A2, GNRH1, APP, HRAS, GPX4, PPP3CA, HMGCR, LMNA, EEF2, ULK3, DAG1, FOXO4, ENDOG, IDE, CDKN1C, KCNE2, NAPEPLD, ASS1, BMPR1A, P2RY1, TP63, RGN, ALOX12, PAWR, SLC30A10, CDKN2B, PTH1R, CAT, PENK, RNF165, EPO, PDCD4, CTSV, DNMBP, NUAK1, TFCP2L1, SNCA, NQO1, PPP1R9A, WNT16, NPY2R, AGT, ALDH3A1, PLA2R1, CLDN1, ADRA1A, PRELP, SLC12A2, GSN, APOD, CRYAB, PPARGC1A, HELT, CYP1A1, KRTAP4-8, KRT25, NPY5R, PCK1, KRTAP4-9, KRT83, LEP, KRT33B, KRTAP4-3</i>
GO:0007568	Envelhecimento	193	3.00E-03	2.78E-02	2.21E-02	

APÊNDICE B – RESULTADO DA ANÁLISE DIFERENCIAL DOS GENES DE INTERESSE POR CATEGORIAS

(continua)

Categoria	Nome	Símbolo Oficial	Presente? ¹	DE? ²	Alteração (log ₂)	Erro padrão (log ₂)	Alteração	Valor p	p ajustado ³	Significancia ⁴
Citotoxicidade	FAS-L	<i>FASLG</i>	Sim	Sim	5.65	0.44	50.29	2.77E-38	2.17E-36	***
	Granzima A	<i>GZMA</i>	Sim	Sim	6.52	0.40	91.97	2.30E-60	1.22E-57	***
	Granzima B	<i>GZMB</i>	Sim	Sim	7.55	0.50	187.98	8.24E-51	1.85E-48	***
	Granzima H	<i>GZMH</i>	Sim	Sim	6.39	0.45	83.69	6.39E-46	9.09E-44	***
	Granzima K	<i>GZMK</i>	Sim	Sim	6.58	0.39	95.37	3.07E-63	2.16E-60	***
	Perforina	<i>PRF1</i>	Sim	Sim	6.16	0.47	71.62	6.75E-39	5.63E-37	***
Controle de proliferação e ciclo celular	Ciclina D3	<i>CCND3</i>	Sim	Sim	1.28	0.15	2.43	3.45E-17	5.98E-16	***
	Ciclina E1	<i>CCNE1</i>	Sim	Sim	1.38	0.20	2.60	1.73E-12	1.89E-11	***
	p21	<i>CDKN1A</i>	Sim	Sim	1.32	0.28	2.50	2.29E-06	1.12E-05	***
	p16	<i>CDKN2A</i>	Sim	Sim	1.96	0.40	3.90	1.12E-06	5.75E-06	***
	p15	<i>CDKN2B</i>	Sim	Sim	-1.49	0.38	-2.81	7.21E-05	2.71E-04	***
Fatores de Transcrição	EOMES	<i>EOMES</i>	Sim	Sim	5.89	0.36	59.17	1.41E-61	8.93E-59	***
	FOXO1	<i>FOXO1</i>	Sim	Sim	-0.38	0.14	-1.30	7.01E-03	1.72E-02	*
	FOXO3	<i>FOXO3</i>	Sim	Não	-0.02	0.14	-1.02	8.64E-01	9.07E-01	
	FOXO3	<i>FOXO3B</i>	Sim	Sim	-0.72	0.30	-1.65	1.64E-02	3.63E-02	*
	FOXO4	<i>FOXO4</i>	Sim	Sim	-0.82	0.14	-1.77	9.06E-09	6.34E-08	***
	FOXP3	<i>FOXP3</i>	Sim	Sim	2.61	0.42	6.09	6.62E-10	5.37E-09	***
	Blimp-1	<i>PRDM1</i>	Sim	Sim	1.57	0.27	2.96	8.69E-09	6.09E-08	***
	RORC	<i>RORC</i>	Sim	Sim	-1.54	0.24	-2.91	1.57E-10	1.37E-09	***
	Tbet	<i>TBX21</i>	Sim	Sim	5.28	0.40	38.97	1.95E-39	1.69E-37	***
	PLZF	<i>ZBTB16</i>	Sim	Sim	-1.74	0.35	-3.34	9.13E-07	4.75E-06	***
	ZEB2	<i>ZEB2</i>	Sim	Sim	1.11	0.18	2.16	2.57E-10	2.19E-09	***

(continuação)

Ligantes de células NK	E-caderina	<i>CDH1</i>	Sim	Não	-0.85	0.39	-1.80	3.12E-02	6.35E-02	
	MHC-I	<i>HLA-E</i>	Sim	Sim	1.33	0.12	2.51	1.90E-27	7.38E-26	***
	MICA	<i>MICA</i>	Sim	Sim	2.05	0.56	4.13	2.62E-04	8.87E-04	***
	MICB	<i>MICB</i>	Sim	Sim	3.82	0.25	14.15	9.35E-54	2.51E-51	***
	DR4/TRAILR1	<i>TNFRSF10A</i>	Sim	Sim	1.09	0.20	2.12	3.22E-08	2.08E-07	***
	DR5/TRAILR2	<i>TNFRSF10B</i>	Sim	Sim	1.48	0.19	2.79	3.50E-15	4.99E-14	***
Moléculas adaptadoras de sinalização	SAP	<i>SH2D1A</i>	Sim	Sim	5.04	0.36	32.87	1.21E-44	1.63E-42	***
	EAT-2	<i>SH2D1B</i>	Sim	Sim	3.75	0.42	13.43	3.96E-19	8.22E-18	***
	SYK	<i>SYK</i>	Sim	Sim	1.75	0.21	3.37	1.86E-17	3.32E-16	***
	DAP12	<i>TYROBP</i>	Sim	Sim	3.59	0.27	12.05	1.29E-39	1.13E-37	***
Receptores coestimulatórios e de <i>homing</i>	CCR7	<i>CCR7</i>	Sim	Sim	4.16	0.37	17.88	5.28E-30	2.47E-28	***
	CD27	<i>CD27</i>	Sim	Sim	3.94	0.30	15.35	2.96E-40	2.73E-38	***
	PD-L1	<i>CD274</i>	Sim	Sim	5.59	0.41	48.16	9.88E-42	1.01E-39	***
	CD28	<i>CD28</i>	Sim	Sim	2.94	0.25	7.70	3.49E-31	1.76E-29	***
	CD69	<i>CD69</i>	Sim	Sim	5.11	0.42	34.56	2.57E-34	1.55E-32	***
	CD62	<i>SELP</i>	Sim	Não	0.62	0.34	1.54	6.59E-02	1.20E-01	
Receptores de células NK	CD57	<i>B3GAT1</i>	Sim	Não	0.28	0.29	1.21	3.36E-01	4.54E-01	
	DNAM-1	<i>CD226</i>	Sim	Sim	4.04	0.29	16.45	6.42E-44	8.03E-42	***
	NKR2B4	<i>CD244</i>	Sim	Sim	4.59	0.38	24.03	1.64E-33	9.35E-32	***
	CD16	<i>FCGR3A</i>	Sim	Sim	6.03	0.32	65.27	6.19E-79	1.24E-75	***
	CD16	<i>FCGR3B</i>	Sim	Sim	3.37	0.38	10.34	2.74E-19	5.76E-18	***
	KIR2DL	<i>KIR2DL1</i>	Sim	Sim	4.07	0.78	16.85	2.05E-07	1.19E-06	***
	KIR2DL	<i>KIR2DL2</i>	Sim	Sim	3.94	1.45	15.36	6.67E-03	1.65E-02	*
	KIR2DL	<i>KIR2DL3</i>	Sim	Sim	6.45	0.98	87.46	4.72E-11	4.40E-10	***
	KIR2DL	<i>KIR2DL4</i>	Sim	Sim	7.67	0.75	203.28	1.03E-24	3.27E-23	***
	KIR2DS	<i>KIR2DS1</i>	Sim	Sim	4.31	1.44	19.79	2.76E-03	7.52E-03	**

(continuação)

	KIR2DS	<i>KIR2DS2</i>	Sim	Sim	5.69	1.87	51.76	2.29E-03	6.36E-03	**
	KIR2DS	<i>KIR2DS3</i>	Sim	Não	4.92	3.42	30.21	1.51E-01	2.39E-01	
	KIR2DS	<i>KIR2DS4</i>	Sim	Sim	4.90	1.28	29.86	1.25E-04	4.51E-04	***
	KIR2DS	<i>KIR2DS5</i>	Sim	Não	3.02	1.62	8.09	6.33E-02	1.16E-01	
	KIR3DL	<i>KIR3DL1</i>	Sim	Sim	3.98	0.84	15.81	1.90E-06	9.41E-06	***
	KIR3DL	<i>KIR3DL2</i>	Sim	Sim	6.76	0.80	108.12	2.52E-17	4.43E-16	***
	KIR3DL	<i>KIR3DL3</i>	Sim	Não	2.01	1.48	4.02	1.73E-01	2.67E-01	
	KIR3DS	<i>KIR3DP1</i>	Sim	Não	2.09	1.07	4.25	5.03E-02	9.54E-02	
	KIR3DS	<i>KIR3DS1</i>	Sim	Não	3.72	2.22	13.17	9.37E-02	1.61E-01	
	CD161	<i>KLRB1</i>	Sim	Sim	3.40	0.35	10.54	2.68E-22	7.08E-21	***
	NKG2A/B	<i>KLRC1</i>	Sim	Sim	4.28	0.47	19.37	9.91E-20	2.17E-18	***
	NKG2C	<i>KLRC2</i>	Sim	Sim	5.79	0.51	55.16	2.37E-29	1.06E-27	***
	NKG2E	<i>KLRC3</i>	Sim	Sim	5.81	0.59	55.98	4.50E-23	1.26E-21	***
	NKG2F	<i>KLRC4</i>	Sim	Sim	3.78	0.45	13.70	2.15E-17	3.79E-16	***
	CD94	<i>KLRD1</i>	Sim	Sim	4.46	0.38	21.98	4.67E-32	2.45E-30	***
	NKp80	<i>KLRF1</i>	Sim	Sim	2.50	0.35	5.65	4.65E-13	5.37E-12	***
	KLRG1	<i>KLRG1</i>	Sim	Sim	2.85	0.36	7.19	3.38E-15	4.84E-14	***
	NKG2D	<i>KLRK1</i>	Sim	Sim	1.93	0.29	3.80	5.98E-11	5.49E-10	***
	CD56	<i>NCAM1</i>	Sim	Sim	1.41	0.36	2.66	7.30E-05	2.74E-04	***
	NKp46	<i>NCR1</i>	Sim	Sim	5.12	0.51	34.88	5.84E-24	1.76E-22	***
	NKp44	<i>NCR2</i>	Sim	Não	0.39	1.60	1.31	8.10E-01	8.68E-01	
	NKp30	<i>NCR3</i>	Sim	Sim	2.71	0.39	6.55	2.81E-12	3.00E-11	***
	TRAIL	<i>TNFSF10</i>	Sim	Sim	2.16	0.20	4.48	1.58E-26	5.77E-25	***
Receptores de citocinas	IL12R	<i>IL12RB1</i>	Sim	Sim	3.49	0.26	11.22	7.13E-42	7.42E-40	***
	IL12R	<i>IL12RB2</i>	Sim	Sim	3.59	0.34	12.04	1.68E-25	5.61E-24	***
	IL15R	<i>IL15RA</i>	Sim	Sim	2.24	0.23	4.73	1.85E-22	4.96E-21	***
	IL18R1	<i>IL18R1</i>	Sim	Sim	1.71	0.19	3.28	1.27E-19	2.75E-18	***
	IL18RAP	<i>IL18RAP</i>	Sim	Sim	5.26	0.41	38.20	5.12E-37	3.67E-35	***

(continuação)

	IL2R	<i>IL2RA</i>	Sim	Sim	5.39	0.38	42.00	3.08E-46	4.60E-44	***
	IL2R	<i>IL2RB</i>	Sim	Sim	3.84	0.41	14.29	7.16E-21	1.69E-19	***
	IL2R	<i>IL2RG</i>	Sim	Sim	4.81	0.35	28.13	3.27E-43	3.83E-41	***
	IL7R	<i>IL7R</i>	Sim	Sim	3.74	0.23	13.38	5.48E-59	2.59E-56	***
SASP	ADAM28	<i>ADAM28</i>	Sim	Sim	3.29	0.26	9.75	8.67E-38	6.54E-36	***
	Angiogenina	<i>ANG</i>	Sim	Sim	-1.50	0.20	-2.83	3.00E-14	3.89E-13	***
	Anfiregulina	<i>AREG</i>	Sim	Sim	2.72	0.54	6.58	3.96E-07	2.18E-06	***
	CCL1	<i>CCL1</i>	Sim	Não	1.13	0.73	2.18	1.23E-01	2.03E-01	
	CCL11	<i>CCL11</i>	Sim	Sim	4.35	0.85	20.45	3.48E-07	1.93E-06	***
	CCL13	<i>CCL13</i>	Sim	Não	-0.71	0.48	-1.63	1.43E-01	2.28E-01	
	CCL16	<i>CCL16</i>	Sim	Não	0.58	0.30	1.49	5.25E-02	9.91E-02	
	CCL2	<i>CCL2</i>	Sim	Sim	2.78	0.31	6.86	8.62E-19	1.74E-17	***
	CCL20	<i>CCL20</i>	Sim	Sim	5.15	0.57	35.59	2.48E-19	5.27E-18	***
	CCL25	<i>CCL25</i>	Sim	Sim	5.13	0.83	35.14	5.38E-10	4.41E-09	***
	CCL26	<i>CCL26</i>	Sim	Não	-0.21	0.49	-1.16	6.65E-01	7.56E-01	
	CCL3	<i>CCL3</i>	Sim	Sim	6.95	0.48	124.03	2.03E-47	3.43E-45	***
	CCL7	<i>CCL7</i>	Sim	Sim	6.52	0.70	91.97	6.91E-21	1.64E-19	***
	CCL8	<i>CCL8</i>	Sim	Sim	7.18	0.46	144.69	6.02E-55	1.84E-52	***
	GM-CSF	<i>CSF2</i>	Sim	Sim	5.56	0.66	47.17	6.11E-17	1.04E-15	***
	G-CSF	<i>CSF3</i>	Sim	Sim	7.72	0.90	210.57	1.08E-17	1.98E-16	***
	Catepsina B	<i>CTSB</i>	Sim	Sim	3.24	0.23	9.42	1.99E-44	2.65E-42	***
	CXCL1	<i>CXCL1</i>	Sim	Sim	7.23	0.60	149.68	1.29E-33	7.37E-32	***
	CXCL11	<i>CXCL11</i>	Sim	Sim	8.63	0.55	395.86	1.22E-55	3.92E-53	***
	CXCL12	<i>CXCL12</i>	Sim	Não	-0.02	0.39	-1.02	9.54E-01	9.69E-01	
	CXCL13	<i>CXCL13</i>	Sim	Sim	8.01	0.73	257.74	6.19E-28	2.51E-26	***
	CXCL2	<i>CXCL2</i>	Sim	Sim	5.07	0.47	33.50	6.72E-27	2.53E-25	***
	CXCL3	<i>CXCL3</i>	Sim	Sim	6.56	0.63	94.51	2.52E-25	8.29E-24	***
	IL-8	<i>CXCL8</i>	Sim	Sim	7.01	0.94	128.60	7.24E-14	9.06E-13	***

(continuação)

EGF	<i>EGF</i>	Sim	Sim	-2.77	0.42	-6.81	5.24E-11	4.84E-10	***
EGFR	<i>EGFR</i>	Sim	Sim	-2.31	0.30	-4.95	2.21E-14	2.89E-13	***
Epiregulina	<i>EREG</i>	Sim	Sim	-0.97	0.34	-1.95	4.82E-03	1.24E-02	*
FasR	<i>FAS</i>	Sim	Sim	1.51	0.15	2.85	9.38E-23	2.58E-21	***
FGF2	<i>FGF2</i>	Sim	Sim	-0.69	0.26	-1.61	9.51E-03	2.26E-02	*
FGF7	<i>FGF7</i>	Sim	Não	0.53	0.45	1.45	2.34E-01	3.41E-01	
Fibronectina	<i>FN1</i>	Sim	Sim	2.91	0.38	7.54	1.35E-14	1.81E-13	***
HGF	<i>HGF</i>	Sim	Sim	2.95	0.30	7.75	4.81E-23	1.34E-21	***
ICAM1	<i>ICAM1</i>	Sim	Sim	4.62	0.36	24.65	1.25E-37	9.36E-36	***
ICAM3	<i>ICAM3</i>	Sim	Sim	2.59	0.21	6.01	3.44E-36	2.28E-34	***
IGFBP2	<i>IGFBP2</i>	Sim	Não	-0.71	0.44	-1.63	1.09E-01	1.83E-01	
IGFBP3	<i>IGFBP3</i>	Sim	Não	0.10	0.27	1.07	7.23E-01	8.02E-01	
IGFBP4	<i>IGFBP4</i>	Sim	Sim	1.30	0.31	2.46	3.26E-05	1.30E-04	***
IGFBP5	<i>IGFBP5</i>	Sim	Não	-0.58	0.37	-1.49	1.18E-01	1.95E-01	
IGFBP6	<i>IGFBP6</i>	Sim	Sim	-2.87	0.50	-7.31	1.28E-08	8.74E-08	***
IGFBP7	<i>IGFBP7</i>	Sim	Sim	0.65	0.23	1.57	4.96E-03	1.27E-02	*
IL-13	<i>IL13</i>	Sim	Sim	2.11	0.62	4.32	6.06E-04	1.91E-03	**
IL-15	<i>IL15</i>	Sim	Sim	2.72	0.23	6.57	5.19E-32	2.72E-30	***
IL-16	<i>IL16</i>	Sim	Sim	1.32	0.17	2.50	9.30E-15	1.27E-13	***
IL-18	<i>IL18</i>	Sim	Sim	-1.29	0.28	-2.44	3.46E-06	1.64E-05	***
IL-1A	<i>IL1A</i>	Sim	Sim	3.76	0.73	13.54	2.33E-07	1.34E-06	***
IL-1β	<i>IL1B</i>	Sim	Sim	6.90	0.72	119.02	9.05E-22	2.31E-20	***
IL-6	<i>IL6</i>	Sim	Sim	7.84	0.63	229.47	4.59E-36	3.02E-34	***
CD130	<i>IL6ST</i>	Sim	Não	-0.13	0.12	-1.09	2.75E-01	3.87E-01	
IL-7	<i>IL7</i>	Sim	Não	0.30	0.18	1.23	8.98E-02	1.56E-01	
MIF	<i>MIF</i>	Sim	Não	0.42	0.19	1.34	2.45E-02	5.16E-02	
MIP	<i>MIP</i>	Sim	Não	0.06	0.29	1.05	8.25E-01	8.80E-01	
MMP1	<i>MMP1</i>	Sim	Sim	9.06	0.61	535.20	2.29E-49	4.51E-47	***
MMP10	<i>MMP10</i>	Sim	Sim	7.38	0.83	166.39	5.09E-19	1.05E-17	***

(continuação)

	MMP12	<i>MMP12</i>	Sim	Sim	5.60	0.90	48.40	5.36E-10	4.40E-09	***
	MMP13	<i>MMP13</i>	Sim	Sim	3.31	0.63	9.89	1.61E-07	9.44E-07	***
	MMP14	<i>MMP14</i>	Sim	Sim	2.08	0.21	4.22	3.45E-22	9.05E-21	***
	MMP2	<i>MMP2</i>	Sim	Sim	1.80	0.26	3.48	2.70E-12	2.88E-11	***
	MMP3	<i>MMP3</i>	Sim	Sim	8.90	0.63	476.63	3.68E-45	5.09E-43	***
	MMP7	<i>MMP7</i>	Sim	Não	-0.37	0.50	-1.29	4.54E-01	5.72E-01	
	MMP9	<i>MMP9</i>	Sim	Sim	4.21	0.50	18.50	3.23E-17	5.63E-16	***
	NGF	<i>NGF</i>	Sim	Não	0.23	0.46	1.18	6.08E-01	7.09E-01	
	PIGF	<i>PIGF</i>	Sim	Não	-0.18	0.12	-1.13	1.44E-01	2.30E-01	
	CD87	<i>PLAUR</i>	Sim	Sim	4.27	0.35	19.30	1.95E-33	1.10E-31	***
	SERPINB2	<i>SERPINB2</i>	Sim	Não	0.24	0.52	1.18	6.43E-01	7.38E-01	
	SERPINE1	<i>SERPINE1</i>	Sim	Sim	4.58	0.34	23.91	3.03E-42	3.24E-40	***
	TIMP1	<i>TIMP1</i>	Sim	Sim	2.71	0.24	6.54	5.10E-29	2.22E-27	***
	TIMP2	<i>TIMP2</i>	Sim	Não	-0.03	0.28	-1.02	9.24E-01	9.49E-01	
	TRAILR3	<i>TNFRSF10C</i>	Sim	Sim	1.83	0.45	3.55	4.97E-05	1.92E-04	***
	TNFRSF11B	<i>TNFRSF11B</i>	Sim	Não	0.63	0.36	1.54	8.30E-02	1.46E-01	
	TNFR1	<i>TNFRSF1A</i>	Sim	Sim	0.37	0.15	1.30	1.20E-02	2.77E-02	*
	TNFR2	<i>TNFRSF1B</i>	Sim	Sim	3.27	0.24	9.67	2.41E-43	2.87E-41	***
Sinalização	ATF	<i>ATF1</i>	Sim	Não	0.12	0.13	1.08	3.56E-01	4.74E-01	
	ATF	<i>ATF2</i>	Sim	Sim	0.33	0.12	1.25	5.83E-03	1.46E-02	*
	ATF	<i>ATF3</i>	Sim	Sim	1.14	0.25	2.20	6.54E-06	2.96E-05	***
	ATF	<i>ATF4</i>	Sim	Sim	0.39	0.14	1.31	5.42E-03	1.37E-02	*
	ATF	<i>ATF5</i>	Sim	Sim	2.51	0.27	5.68	6.89E-21	1.63E-19	***
	ATF	<i>ATF6</i>	Sim	Sim	0.38	0.10	1.31	7.84E-05	2.93E-04	***
	ATF	<i>ATF7</i>	Sim	Não	-0.14	0.15	-1.10	3.28E-01	4.46E-01	
	ATM	<i>ATM</i>	Sim	Sim	1.04	0.18	2.06	3.58E-09	2.65E-08	***
	p38	<i>MAPK14</i>	Sim	Não	-0.15	0.09	-1.11	8.72E-02	1.52E-01	
	p-AMPK	<i>PRKAA1</i>	Sim	Sim	0.28	0.08	1.22	6.29E-04	1.97E-03	**

(conclusão)										
	p-AMPK	<i>PRKAA2</i>	Sim	Sim	-2.68	0.26	-6.43	1.96E-25	6.47E-24	***
	p-AMPK	<i>PRKAB1</i>	Sim	Não	0.18	0.09	1.13	3.47E-02	6.96E-02	
	p-AMPK	<i>PRKAB2</i>	Sim	Sim	-0.75	0.17	-1.68	1.10E-05	4.77E-05	***
	p-AMPK	<i>PRKAG2</i>	Sim	Sim	0.42	0.14	1.34	3.82E-03	1.01E-02	*
	Sestrin 1	<i>SESN1</i>	Sim	Sim	-1.18	0.16	-2.27	6.62E-14	8.30E-13	***
	Sestrin 2	<i>SESN2</i>	Sim	Sim	1.08	0.23	2.12	2.90E-06	1.40E-05	***
	Sestrin 3	<i>SESN3</i>	Sim	Sim	-1.00	0.19	-2.00	1.11E-07	6.69E-07	***
	JNK	<i>MAPK8</i>	Sim	Não	0.18	0.13	1.13	1.59E-01	2.49E-01	
	ERK	<i>MAPK1</i>	Sim	Não	0.03	0.10	1.02	7.58E-01	8.29E-01	
Transdução de sinal em linfócitos T	Complexo CD3	<i>CD3D</i>	Sim	Sim	5.15	0.36	35.43	1.01E-46	1.58E-44	***
	Complexo CD3	<i>CD3E</i>	Sim	Sim	4.35	0.32	20.34	2.21E-41	2.19E-39	***
	Complexo CD3	<i>CD3G</i>	Sim	Sim	4.57	0.29	23.76	1.71E-55	5.43E-53	***
	DAP10	<i>HCST</i>	Sim	Sim	3.76	0.31	13.59	3.69E-34	2.20E-32	***
	LAT	<i>LAT</i>	Sim	Sim	3.31	0.30	9.90	6.48E-29	2.79E-27	***
	LCK	<i>LCK</i>	Sim	Sim	4.44	0.31	21.77	4.12E-47	6.75E-45	***
	PLCG1	<i>PLCG1</i>	Sim	Sim	0.44	0.15	1.36	3.20E-03	8.58E-03	**
	TCR- α	<i>TRA</i>	Não	Não						
	ZAP70	<i>ZAP70</i>	Sim	Sim	4.71	0.37	26.23	1.80E-36	1.23E-34	***

¹ = Se o gene está presente no conjunto de dados analisado.

² = Se o gene está diferencialmente expresso entre os grupos.

³ = Valor de p ajustado pelo método de Benjamini-Hochberg

⁴ = Significância estatística (* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$)

APÊNDICE C – RESULTADO DA ANÁLISE DE DECONVOLUÇÃO CELULAR DOS INDIVÍDUOS CONTROLE

(continua)

Cell Subtype	Relative Scores - Controls						
	HS1	HS2	HS3	HS4	HS5	HS6	HS7
Basophils	0.00	0.00	0.10	0.00	-0.08	0.00	0.00
CD4 ⁺ Central Memory	0.00	0.00	0.00	0.00	-0.12	0.04	0.00
CD4 ⁺ Effector Memory	0.00	0.00	0.00	0.00	0.00	0.01	0.00
CD8 ⁺ Central Memory	0.00	0.00	0.00	0.00	-0.03	0.00	0.00
CD8 ⁺ Effector Memory	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CD8 ⁺ Effector Memory RA	0.00	0.00	-0.09	0.00	0.16	0.00	0.00
Colony Forming Unit-Granulocyte	0.00	0.00	-0.02	0.00	0.00	0.12	0.00
Colony Forming Unit-Megakaryocytic	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Colony Forming Unit-Monocyte	-0.07	0.00	-0.10	0.00	0.01	0.10	0.00
Common myeloid progenitor	0.00	0.00	0.00	0.00	0.00	0.06	0.00
Early B-cell	0.00	-0.02	0.09	-0.01	0.00	0.00	0.05
Eosinophill	-0.17	0.00	0.00	0.00	-0.11	0.04	0.00
Erythroid_CD34 ⁻ CD71 ⁻ GlyA ⁺	0.00	-0.07	0.00	0.00	0.00	-0.01	0.00
Erythroid_CD34 ⁻ CD71 ⁺ GlyA ⁻	0.17	0.00	-0.09	-0.06	0.00	0.00	0.00
Erythroid_CD34 ⁻ CD71 ⁺ GlyA ⁺	0.01	0.00	0.00	0.00	0.00	-0.03	0.00
Erythroid_CD34 ⁻ CD71 ^{lo} GlyA ⁺	0.03	-0.02	0.00	0.00	0.06	-0.03	0.00
Erythroid_CD34 ⁺ CD71 ⁺ GlyA ⁻	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Granulocyte (Neutrophil)	0.00	0.05	-0.04	0.00	0.00	0.08	-0.05
Granulocyte (Neutrophilic Metamyelocyte)	-0.15	0.07	-0.03	0.12	0.00	0.10	-0.12
Granulocyte/monocyte progenitor	-0.17	0.01	0.00	0.00	0.00	0.15	0.00
Hematopoietic stem cell_CD133 ⁺ CD34 ^{dim}	0.00	0.00	0.00	0.00	0.00	-0.13	0.00
Hematopoietic stem cell_CD38 ⁻ CD34 ⁺	0.00	0.04	0.00	0.02	0.00	-0.18	0.00
Mature B-cell class able to switch	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mature B-cell class switched	0.00	0.00	0.00	0.00	0.01	0.00	0.00
Mature B-cells	0.00	0.02	0.02	0.00	-0.06	-0.07	0.02
Mature NK cell_CD56 ⁻ CD16 ⁻ CD3 ⁻	0.00	0.00	0.00	-0.08	0.00	0.00	0.00
Mature NK cell_CD56 ⁻ CD16 ⁺ CD3 ⁻	-0.02	0.00	0.00	0.00	0.06	0.00	0.00
Mature NK cell_CD56 ⁺ CD16 ⁺ CD3 ⁻	0.00	0.00	0.00	-0.01	0.00	0.00	0.00
Megakaryocyte	0.00	0.00	0.00	0.04	-0.04	0.00	0.00

							(conclusão)
Megakaryocyte/ erythroid progenitor	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Monocyte	-0.01	0.01	-0.20	0.08	0.20	0.03	-0.10
Myeloid Dendritic Cell	-0.02	0.12	0.00	0.00	0.00	0.00	-0.10
Naive B-cells	0.00	0.00	0.00	0.00	0.00	-0.04	0.00
Naive CD4 ⁺ T-cell	0.00	-0.04	0.00	0.00	-0.04	0.00	0.00
Naive CD8 ⁺ T-cell	0.00	0.00	0.00	0.03	0.00	0.00	0.00
NKT	0.00	0.00	0.00	0.00	-0.04	0.01	0.00
Plasmacytoid Dendritic Cell	0.00	0.00	0.00	-0.06	0.00	0.00	0.00
Pro B-cell	0.00	-0.09	0.00	-0.06	0.00	0.00	0.05

APÊNDICE E – SCRIPT DA ANÁLISE E INFORMAÇÕES DE SESSÃO DO R

```

## Author of this script: Carlos Henrique Fantecelle
##
## Analysis of the Gene Expression data (RNA-Seq) from Amorim et al (2019):
## Variable gene
## expression and parasite load predict treatment outcome in cutaneous
## leishmaniasis.
##
## GSE127831, available at Gene Expression Omnibus. The data was downloaded as a
## txt tab delimited file as raw, non normalized, data.
##
## Last Update: 29/07/2020

##----- Working Directory -----
setwd("C:/Users/Laboratorio/OneDrive/Analises_LC/")

##----- Dependencies -----

## Data Import, Manipulation & Visualisation
library(readxl)
library(tidyverse)
library(pheatmap)
library(RColorBrewer)
library(circlize)
library(dendextend)
library(xlsx)
library(reshape2)

## Bioconductor Packages
library(GEOquery)
library(Biobase)
library(GenomicFeatures)
library(BiocParallel)
library(DESeq2)
library(vsn)
library(biomaRt)
library(GO.db)
library(EnsDb.Hsapiens.v86)
library(clusterProfiler)
library(enrichplot)
library(org.Hs.eg.db)
library(ComplexHeatmap)
library(ggcorrplot)
library(EnhancedVolcano)

## Useful function
# `notin` <- Negate(`in`)

##----- Importing Data -----

## Getting the ExpressionSet from GEO
rnaseq_eset <- getGEO("GSE127831")
rnaseq_eset <- rnaseq_eset[[1]]
rnaseq_eset

## AssayData is empty (0 features), so data was downloaded as raw, unnormalized,
counts
## from GEO website (DESeq2 requires unnormalized data).

count_data <- read_delim("GSE127831_Amorim_GEO_raw.txt", "\t", escape_double =
FALSE,
                        trim_ws = TRUE)

## Converting to data frame
count_data <- as.data.frame(count_data)

## Assigning genes as row names
rownames(count_data) <- count_data$X1

## Deleting extra column
count_data <- count_data[,-1]

```

```

## Getting phenotypic data from eSet
pheno_data <- pData(rnaseq_eset)

##----- Cleaning & Tidying phenotypic data -----
pheno_data <- as.data.frame(pheno_data)

## Storing unused data
unused_pdata <- pheno_data[, c(3:7, 9, 21:58)]

## Removing unused data
pheno_data <- pheno_data[, -c(3:7, 9, 21:58)]

## Removing duplicated data
pheno_data <- pheno_data[, -c(4:14)]

## Adjusting colnames
colnames(pheno_data) <- c("sample_name", "geo_acc", "source", "age", "dth_mm3",
"infection",
"lesion_size_mm3", "n_parasite_tx_rnaseq",
"n_parasites_qpcr",
"sex", "subject", "days_cure", "tissue",
"treatment_outcome")

## Shifting columns
pheno_data <- pheno_data[, c(1:3, 6, 11, 13, 14, 4:5, 7:10, 12)]

## Removing redundant columns
pheno_data <- pheno_data[, -c(4:6)]

## Adjusting factors
pheno_data$source <- str_replace_all(pheno_data$source, "healthy skin",
"healthy_skin")
pheno_data$source <- str_replace_all(pheno_data$source, "cutaneous leishmaniasis
lesion",
"CL_lesion")

pheno_data[pheno_data == "na"] <- NA

## Correcting sample names to match counts matrix
pheno_data$sample_name <- str_replace_all(pheno_data$sample_name, "skin_", "")

## Checking if they match
identical(colnames(count_data), pheno_data$sample_name) # True

## Adjusting rownames from pheno_data
rownames(pheno_data) <- pheno_data$sample_name

## DESeq2 requires integers as counts, while the software used by the original
authors works
## with an estimate. Michael Love, author of DESeq, stated that rounding the counts
should not
## lead to loss of precision, since fractions of these counts are very small
compared to bio-
## logical variation on RNA-Seq counts.
[https://support.bioconductor.org/p/88763/#88764]

## Rounding counts for use in DESeq2
round_counts <- round(count_data)

## Releveling factors
pheno_data$source <- as.factor(pheno_data$source)
levels(pheno_data$source)
pheno_data$source <- relevel(pheno_data$source, "healthy_skin")
levels(pheno_data$source)

##----- Differential Expression Analysis using DESeq2 -----
## Creating DESeq object
CL_dds <- DESeqDataSetFromMatrix(countData = round_counts, colData = pheno_data,
design = ~ source)

## Filtering low count genes
keep <- rowSums(counts(CL_dds)) >= 10
CL_dds <- CL_dds[keep,]
CL_dds

```

```

## DESeq2
dds <- DESeq(CL_dds)

## Extracting results
res <- results(dds, alpha = 0.05)
res
summary(res)

## Extracting res object from genes of interest
res_total_int <- subset(res, rownames(res) %in% genes_int$hgnc_symbol)

# Subsetting res to get only significant genes
res_signif <- subset(res, padj < 0.05)
summary(res_signif)

## Filtered Genes analysis
metadata(res)$filterThreshold

plot(metadata(res)$filterNumRej, type = "b", ylab = "number of rejections",
      xlab = "quantiles of filter")
lines(metadata(res)$lo.fit, col = "red")
abline(v = metadata(res)$filterTheta)

## MA-plot to see dispersion
plotMA(res, ylim = c(-2, 2), alpha = 0.05)

##----- Count data transformations -----
rld <- rlog(dds, blind = FALSE)
ntd <- normTransform(dds)
vst <- vst(dds, blind = FALSE)

## Checking effect on the variance
meanSdPlot(assay(ntd))
meanSdPlot(assay(rld))
meanSdPlot(assay(vst)) # more homoskedastic

## Writing table for use in immquant software
write.table(assay(vst), "vst_normalized_counts_immquant.txt", sep="\t", quote =
FALSE)

##----- Quality assessment by sample clustering and visualization -----
## Heatmaps of counts
df <- as.data.frame(colData(dds)[,"source"])
rownames(df) <- colData(dds)$sample_name
select <- order(rowMeans(counts(dds, normalized = TRUE)), decreasing = TRUE)[1:20]

hm_ntd <- pheatmap(assay(ntd)[select,], cluster_rows = FALSE, show_rownames =
FALSE,
                  cluster_cols = FALSE, annotation_col = df, show_colnames =
FALSE)

hm_rld <- pheatmap(assay(rld)[select,], cluster_rows = FALSE, show_rownames =
FALSE,
                  cluster_cols = FALSE, annotation_col = df, show_colnames =
FALSE)

## Heatmap of sample-to-sample distances
sampleDists <- dist(t(assay(rld)))

sampleDistMatrix <- as.matrix(sampleDists)
rownames(sampleDistMatrix) <- paste(rld$source)
colnames(sampleDistMatrix) <- NULL
colors <- colorRampPalette(rev(brewer.pal(9, "Blues")))(255)

pheatmap(sampleDistMatrix, clustering_distance_rows = sampleDists,
         clustering_distance_cols = sampleDists, col = colors)

## PCA
plotPCA(vst, intgroup = "source")

## Dispersion plot

```

```

plotDispEsts(dds)

##----- Extracting genes of interest -----

# Importing table of genes of interest
genes_int <- as.data.frame(read_xlsx("checking_genes/genes_total.xlsx"))

## Checking genes of interest in results object

# Setting new columns
genes_int$is_in_dataset <- NA
genes_int$is_DE <- NA

## Gene is on dataset?
for (i in 1:nrow(genes_int)) {
  if (genes_int[i,2] %in% rownames(round_counts)) {
    genes_int[i,6] <- TRUE
  } else {
    genes_int[i,6] <- FALSE
  }
}

## Is the gene a DEG?
for (i in 1:nrow(genes_int)) {
  if (genes_int[i,2] %in% res_signif@rownames) {
    genes_int[i,7] <- TRUE
  } else {
    genes_int[i,7] <- FALSE
  }
}

## Checking numbers
table(genes_int$is_in_dataset)
table(genes_int$is_DE)

## Writing table of genes checked
write_xlsx(genes_int, "checking_genes/genes_total_checked.xlsx", "DE check",
col.names = TRUE,
row.names = TRUE)

# Subsetting DEGs
DGE_check_true <- subset(genes_int, is_DE == TRUE)
DGE_check_false <- subset(genes_int, is_DE == FALSE)

## Plotting specific genes for testing

# Is it a significant DEG?
"KLRG1" %in% DGE_check_true$hgnc_symbol
"SESN2" %in% DGE_check_true$hgnc_symbol
"TYROBP" %in% DGE_check_true$hgnc_symbol
"MICA" %in% DGE_check_true$hgnc_symbol
"MICB" %in% DGE_check_true$hgnc_symbol
"KLRK1" %in% DGE_check_true$hgnc_symbol
"KLRC2" %in% DGE_check_true$hgnc_symbol
"ADAM28" %in% DGE_check_true$hgnc_symbol

# Plot
par(mfrow = c(2, 4))
plotCounts(dds, gene="KLRG1", intgroup="source")
plotCounts(dds, gene="SESN2", intgroup="source")
plotCounts(dds, gene="TYROBP", intgroup="source")
plotCounts(dds, gene="MICA", intgroup="source")
plotCounts(dds, gene="MICB", intgroup="source")
plotCounts(dds, gene="KLRK1", intgroup="source")
plotCounts(dds, gene="KLRC2", intgroup="source")
plotCounts(dds, gene="ADAM28", intgroup="source")

# Creating vector of DEGs and not DEGs
genes_list_true <- as.vector(DGE_check_true$hgnc_symbol)
genes_list_false <- as.vector(DGE_check_false$hgnc_symbol)

##----- Genes of Interest Analysis - Heatmap -----

# Creating neu object to facilitate use and maintain backup
genes_of_interest <- genes_int

```

```

## Subsetting count data by DEGs of interest
round_counts_intDEG <- subset(round_counts,
                             rownames(round_counts) %in%
genes_of_interest$hgnc_symbol[genes_of_interest$is_DE=="TRUE"])

true_DEGs_test <- genes_of_interest$hgnc_symbol[genes_of_interest$is_DE=="TRUE"]
length(true_DEGs_test)

identical(rownames(round_counts_intDEG)[order(rownames(round_counts_intDEG))],
          true_DEGs_test[order(true_DEGs_test)])

dim(round_counts_intDEG)
dim(genes_of_interest)
table(genes_of_interest$is_DE)

# Subsetting dds and vst objects with only DEGs of interest
dds_intDEG <- subset(dds, rownames(dds) %in%
genes_of_interest$hgnc_symbol[genes_of_interest$is_DE=="TRUE"])
vst_intDEG <- subset(vst, rownames(vst) %in%
genes_of_interest$hgnc_symbol[genes_of_interest$is_DE=="TRUE"])

## Heatmap of DEGs

# Centering the samples by rowMeans using vst normalized object
matDEG <- assay(vst_intDEG)
matDEG <- matDEG - rowMeans(matDEG)

## Heatmaps using ComplexHeatmap

# Creating color function
library(circlize)
col_fun <- colorRamp2(c(-5, 0, 5), c("cyan", "black", "yellow"))
max(matDEG_cat)
min(matDEG_cat)

# Adjusting annotations
comp_annot_deg <- data.frame(annot_DEG$Grupo)
colnames(comp_annot_deg) <- "Grupo"
comp_annot_deg$Grupo <- str_replace_all(comp_annot_deg$Grupo, "healthy_skin",
"Healthy skin")
comp_annot_deg$Grupo <- str_replace_all(comp_annot_deg$Grupo, "Healthy skin", "CS")
comp_annot_deg$Grupo <- str_replace_all(comp_annot_deg$Grupo, "CL_lesion", "CL
Lesion")
comp_annot_deg$Grupo <- str_replace_all(comp_annot_deg$Grupo, "CL Lesion", "LCL")

# Defining annotation colors
col_vector <- c("#33C4FF", "#FD4646")
names(col_vector) <- c("CS", "LCL")

# Creating heatmap annotation object
top_annot <- HeatmapAnnotation(df = comp_annot_deg, show_annotation_name = FALSE,
                             col = list(Grupo = col_vector),
                             simple_anno_size = unit(0.2, "cm"))

# Clustering and rotation adjustment of dendrogram
col_dend <- as.dendrogram(hclust(dist(t(matDEG))))
col_dend <- click_rotate(col_dend)

# Inicial heatmap
Heatmap(matDEG, name = "Expressão",
        col = col_fun,
        clustering_distance_rows = "euclidean",
        row_names_gp = gpar(fontsize = 6),
        column_names_gp = gpar(fontsize = 6),
        top_annotation = top_annot,
        cluster_columns = col_dend)

## Creating a Heatmap separating by gene category

# Sorting dataset to match categories
int_DEG_only <- genes_of_interest[genes_of_interest$is_DE=="TRUE",]
int_DEG_only <- int_DEG_only[order(int_DEG_only$hgnc_symbol),]

identical(rownames(matDEG), int_DEG_only$hgnc_symbol)

```

```

matDEG_cat <- as.data.frame(matDEG)
matDEG_cat$categoria <- int_DEG_only$categoria
matDEG_cat$genes <- rownames(matDEG_cat)
matDEG_cat <- arrange(matDEG_cat, matDEG_cat$categoria)
head(matDEG_cat)

DEG_categories <- matDEG_cat[,29:30]
rownames(matDEG_cat) <- matDEG_cat$genes

matDEG_cat$categoria <- NULL
matDEG_cat$genes <- NULL

matDEG_cat <- as.matrix(matDEG_cat)

# Making the heatmap
hm_cat <- Heatmap(matDEG_cat, name = "Expressão",
  col = col_fun,
  clustering_distance_rows = "euclidean",
  row_names_gp = gpar(fontsize = 5, fontface = "italic"),
  column_names_gp = gpar(fontsize = 6),
  top_annotation = top_annot,
  cluster_columns = col_dend,
  row_split = DEG_categories$categoria,
  row_title_rot = 0,
  row_dend_reorder = FALSE,
  row_title_gp = gpar(fontsize = 10, fontface = "bold"),
  show_parent_dend_line = FALSE,
  heatmap_legend_param = list(direction = "horizontal",
  title_position = "topcenter",
  at = c(-5, -2.5, 0, 2.5, 5),
  labels = c("-5", " ", " ", "0", " ", " ",
"5"))))

# Saving the heatmap in high quality
tiff("results/dissertacao_att/hqplots/Heatmap_VST_Categories.tiff", units="in",
width=9, height=11, res=300)

draw(hm_cat, heatmap_legend_side = "right", annotation_legend_side = "bottom")

dev.off()

# Saving the clustering in high quality
tiff("results/dissertacao_att/hqplots/Clustering.tiff", units="in", width=6,
height=4, res=300)

plot(col_dend)

dev.off()

##---- Heatmaps for each individual category -----

# Correcting common names in genes_int_deg
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "Gramzyme",
"Granzima")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "Cyclin",
"Ciclina")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "Perforin",
"Perforina")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot,
"\\(CDKN1A\\)", "\\(p21\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot,
"\\(CDKN2B\\)", "\\(p15\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(ANG\\)",
"\\(Angiogenina\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(AREG\\)",
"\\(Anfiregulina\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(CSF2\\)",
"\\(GM-CSF\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(CSF3\\)",
"\\(G-CSF\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(CTSBB\\)",
"\\(Catepsina B\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(IL8\\)",
"\\(IL-8\\)")

```

```

genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(IL6\\)",
"\\(IL-6\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(IL13\\)",
"\\(IL-13\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(IL15\\)",
"\\(IL-15\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(EREG\\)",
"\\(Epi regulina\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(FASR\\)",
"\\(FasR\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(FN1\\)",
"\\(Fibronectina\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(IL1A\\)",
"\\(IL-1A\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot,
"\\(IL1beta\\)", "\\(IL-1??\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot,
"\\(PLAUR\\)", "\\(CD87\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot,
"\\(TNFRSF10C\\)", "\\(TRAILR3\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot,
"\\(TNFRSF1A\\)", "\\(TNFR1\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot,
"\\(TNFRSF1B\\)", "\\(TNFR2\\)")
genes_int_deg$names_plot <- str_replace_all(genes_int_deg$names_plot, "\\(CD3
complex\\)", "\\(Complexo CD3\\)")

# Separating by categories again
identical(rownames(matDEG), genes_int_deg$hgnc_symbol)
matDEG_cat2 <- as.data.frame(matDEG)
matDEG_cat2$categoria <- genes_int_deg$categoria
matDEG_cat2$genes <- rownames(matDEG_cat2)
matDEG_cat2$names_plot <- genes_int_deg$names_plot
matDEG_cat2 <- arrange(matDEG_cat2, matDEG_cat2$categoria)
head(matDEG_cat2)

# Citotoxicity Genes
mat_citotox <- matDEG_cat2[matDEG_cat2$categoria=="Citotoxicidade",]
rownames(mat_citotox) <- mat_citotox$names_plot
mat_citotox <- as.matrix(mat_citotox[,1:28])

# Signalling Genes
mat_signalling <- matDEG_cat2[matDEG_cat2$categoria=="Sinalização",]
rownames(mat_signalling) <- mat_signalling$names_plot
mat_signalling <- as.matrix(mat_signalling[,1:28])

# NK ligands Genes
mat_NKligands <- matDEG_cat2[matDEG_cat2$categoria=="Ligantes de células NK",]
rownames(mat_NKligands) <- mat_NKligands$names_plot
mat_NKligands <- as.matrix(mat_NKligands[,1:28])

# Proliferation and cell cycle control
mat_prolif <- matDEG_cat2[matDEG_cat2$categoria=="Controle de proliferação e ciclo
celular",]
rownames(mat_prolif) <- mat_prolif$names_plot
mat_prolif <- as.matrix(mat_prolif[,1:28])

# Transcription factors
mat_transc <- matDEG_cat2[matDEG_cat2$categoria=="Fatores de Transcrição",]
rownames(mat_transc) <- mat_transc$names_plot
mat_transc <- as.matrix(mat_transc[,1:28])

# Costimulatory and homing receptors
mat_coest <- matDEG_cat2[matDEG_cat2$categoria=="Receptores coestimulatórios e de
homing",]
rownames(mat_coest) <- mat_coest$names_plot
mat_coest <- as.matrix(mat_coest[,1:28])

# Innate signalling
mat_innate <- matDEG_cat2[matDEG_cat2$categoria=="Moléculas de sinalização inata",]
rownames(mat_innate) <- mat_innate$names_plot
mat_innate <- as.matrix(mat_innate[,1:28])

# Cytokine receptors
mat_cytokine <- matDEG_cat2[matDEG_cat2$categoria=="Receptores de citocinas",]
rownames(mat_cytokine) <- mat_cytokine$names_plot
mat_cytokine <- as.matrix(mat_cytokine[,1:28])

```

```

# T cell signal transduction
mat_tcell <- matDEG_cat2[matDEG_cat2$categoria=="Transdução de sinal em linfócitos
T",]
rownames(mat_tcell) <- mat_tcell$names_plot
mat_tcell <- as.matrix(mat_tcell[,1:28])

# NK Receptors
mat_NKreceptors <- matDEG_cat2[matDEG_cat2$categoria=="receptores de células NK",]
rownames(mat_NKreceptors) <- mat_NKreceptors$names_plot
mat_NKreceptors <- as.matrix(mat_NKreceptors[,1:28])

# NK Receptors
mat_SASP <- matDEG_cat2[matDEG_cat2$categoria=="SASP",]
rownames(mat_SASP) <- mat_SASP$names_plot
mat_SASP <- as.matrix(mat_SASP[,1:28])

## Drawing heatmaps

# NK ligands
hm_NKligands <- Heatmap(mat_NKligands, name = "Expressão",
  col = col_fun,
  column_title = "Ligantes de células NK",
  column_title_gp = gpar(fontsize = 12, fontface = "bold"),
  clustering_distance_rows = "euclidean",
  row_names_gp = gpar(fontsize = 10, fontface = "italic"),
  column_names_gp = gpar(fontsize = 10),
  top_annotation = top_annot,
  cluster_columns = col_dend,
  row_title_rot = 0,
  row_dend_reorder = FALSE,
  row_title_gp = gpar(fontsize = 10, fontface = "bold"),
  show_parent_dend_line = FALSE,
  heatmap_legend_param = list(direction = "horizontal",
title_position = "topcenter",
                                at = c(-5, -2.5, 0, 2.5, 5),
                                labels = c("-5", " ", " ", "0", " ",
"5"))))

draw(hm_NKligands, heatmap_legend_side = "right", annotation_legend_side =
"bottom")

# NK receptors
hm_NKreceptors <- Heatmap(mat_NKreceptors, name = "Expressão",
  col = col_fun,
  column_title = "Receptores de células NK",
  column_title_gp = gpar(fontsize = 12, fontface = "bold"),
  clustering_distance_rows = "euclidean",
  row_names_gp = gpar(fontsize = 10, fontface = "italic"),
  column_names_gp = gpar(fontsize = 10),
  top_annotation = top_annot,
  cluster_columns = col_dend,
  row_title_rot = 0,
  row_dend_reorder = FALSE,
  row_title_gp = gpar(fontsize = 10, fontface = "bold"),
  show_parent_dend_line = FALSE,
  heatmap_legend_param = list(direction = "horizontal",
title_position = "topcenter",
                                at = c(-5, -2.5, 0, 2.5, 5),
                                labels = c("-5", " ", " ", "0", " ",
", "5"))))

draw(hm_NKreceptors, heatmap_legend_side = "right", annotation_legend_side =
"bottom")

# SASP
hm_SASP <- Heatmap(mat_SASP, name = "Expressão",
  col = col_fun,
  column_title = "SASP",
  column_title_gp = gpar(fontsize = 12, fontface = "bold"),
  clustering_distance_rows = "euclidean",
  row_names_gp = gpar(fontsize = 10, fontface = "italic"),
  column_names_gp = gpar(fontsize = 10),
  top_annotation = top_annot,
  cluster_columns = col_dend,
  row_title_rot = 0,
  row_dend_reorder = FALSE,

```

```

        row_title_gp = gpar(fontsize = 10, fontface = "bold"),
        show_parent_dend_line = FALSE,
        heatmap_legend_param = list(direction = "horizontal",
                                     at = c(-5, -2.5, 0, 2.5, 5),
                                     labels = c("-5", " ", "0", " ")),
    title_position = "topcenter",
    ", "5"))))
draw(hm_SASP, heatmap_legend_side = "right", annotation_legend_side = "bottom")

# Signalling
hm_signalling <- Heatmap(mat_signalling, name = "Expressão",
                        col = col_fun,
                        column_title = "Sinalização",
                        column_title_gp = gpar(fontsize = 12, fontface = "bold"),
                        clustering_distance_rows = "euclidean",
                        row_names_gp = gpar(fontsize = 10, fontface = "italic"),
                        column_names_gp = gpar(fontsize = 10),
                        top_annotation = top_annot,
                        cluster_columns = col_dend,
                        row_title_rot = 0,
                        row_dend_reorder = FALSE,
                        row_title_gp = gpar(fontsize = 10, fontface = "bold"),
                        show_parent_dend_line = FALSE,
                        heatmap_legend_param = list(direction = "horizontal",
                                                    at = c(-5, -2.5, 0, 2.5, 5),
                                                    labels = c("-5", " ", "0", " ")),
    title_position = "topcenter",
    "5"))))
draw(hm_signalling, heatmap_legend_side = "right", annotation_legend_side =
"bottom")

# Proliferation and cell cycle control
hm_prolif <- Heatmap(mat_prolif, name = "Expressão",
                    col = col_fun,
                    column_title = "Controle de proliferação e ciclo celular",
                    column_title_gp = gpar(fontsize = 12, fontface = "bold"),
                    clustering_distance_rows = "euclidean",
                    row_names_gp = gpar(fontsize = 10, fontface = "italic"),
                    column_names_gp = gpar(fontsize = 10),
                    top_annotation = top_annot,
                    cluster_columns = col_dend,
                    row_title_rot = 0,
                    row_dend_reorder = FALSE,
                    row_title_gp = gpar(fontsize = 10, fontface = "bold"),
                    show_parent_dend_line = FALSE,
                    heatmap_legend_param = list(direction = "horizontal",
                                                at = c(-5, -2.5, 0, 2.5, 5),
                                                labels = c("-5", " ", "0", " ")),
    title_position = "topcenter",
    ", "5"))))
draw(hm_prolif, heatmap_legend_side = "right", annotation_legend_side = "bottom")

# Transcription factors
hm_transc <- Heatmap(mat_transc, name = "Expressão",
                    col = col_fun,
                    column_title = "Fatores de Transcrição",
                    column_title_gp = gpar(fontsize = 12, fontface = "bold"),
                    clustering_distance_rows = "euclidean",
                    row_names_gp = gpar(fontsize = 10, fontface = "italic"),
                    column_names_gp = gpar(fontsize = 10),
                    top_annotation = top_annot,
                    cluster_columns = col_dend,
                    row_title_rot = 0,
                    row_dend_reorder = FALSE,
                    row_title_gp = gpar(fontsize = 10, fontface = "bold"),
                    show_parent_dend_line = FALSE,
                    heatmap_legend_param = list(direction = "horizontal",
                                                at = c(-5, -2.5, 0, 2.5, 5),
                                                labels = c("-5", " ", "0", " ")),
    title_position = "topcenter",
    "5"))))
draw(hm_transc, heatmap_legend_side = "right", annotation_legend_side = "bottom")

```

```

# Cytotoxicity
hm_citotox <- Heatmap(mat_citotox, name = "Expressão",
  col = col_fun,
  column_title = "Citotoxicidade",
  column_title_gp = gpar(fontsize = 12, fontface = "bold"),
  clustering_distance_rows = "euclidean",
  row_names_gp = gpar(fontsize = 10, fontface = "italic"),
  column_names_gp = gpar(fontsize = 10),
  top_annotation = top_annot,
  cluster_columns = col_dend,
  row_title_rot = 0,
  row_dend_reorder = FALSE,
  row_title_gp = gpar(fontsize = 10, fontface = "bold"),
  show_parent_dend_line = FALSE,
  heatmap_legend_param = list(direction = "horizontal",
    at = c(-5, -2.5, 0, 2.5, 5),
    labels = c("-5", " ", "0", " ",
"5")))
draw(hm_citotox, heatmap_legend_side = "right", annotation_legend_side = "bottom")

# Costimulatory and Homing receptors
hm_coest <- Heatmap(mat_coest, name = "Expressão",
  col = col_fun,
  column_title = "Moléculas costimulatórias e de homing",
  column_title_gp = gpar(fontsize = 12, fontface = "bold"),
  clustering_distance_rows = "euclidean",
  row_names_gp = gpar(fontsize = 10, fontface = "italic"),
  column_names_gp = gpar(fontsize = 10),
  top_annotation = top_annot,
  cluster_columns = col_dend,
  row_title_rot = 0,
  row_dend_reorder = FALSE,
  row_title_gp = gpar(fontsize = 10, fontface = "bold"),
  show_parent_dend_line = FALSE,
  heatmap_legend_param = list(direction = "horizontal",
    at = c(-5, -2.5, 0, 2.5, 5),
    labels = c("-5", " ", "0", " ",
"5")))
draw(hm_coest, heatmap_legend_side = "right", annotation_legend_side = "bottom")

# Signalling adaptor molecules
hm_innate <- Heatmap(mat_innate, name = "Expressão",
  col = col_fun,
  column_title = "Moléculas adaptadoras de sinalização",
  column_title_gp = gpar(fontsize = 12, fontface = "bold"),
  clustering_distance_rows = "euclidean",
  row_names_gp = gpar(fontsize = 10, fontface = "italic"),
  column_names_gp = gpar(fontsize = 10),
  top_annotation = top_annot,
  cluster_columns = col_dend,
  row_title_rot = 0,
  row_dend_reorder = FALSE,
  row_title_gp = gpar(fontsize = 10, fontface = "bold"),
  show_parent_dend_line = FALSE,
  heatmap_legend_param = list(direction = "horizontal",
    at = c(-5, -2.5, 0, 2.5, 5),
    labels = c("-5", " ", "0", " ",
"5")))
draw(hm_innate, heatmap_legend_side = "right", annotation_legend_side = "bottom")

# Cytokine receptors
hm_cytokine <- Heatmap(mat_cytokine, name = "Expressão",
  col = col_fun,
  column_title = "Receptores de citocinas",
  column_title_gp = gpar(fontsize = 12, fontface = "bold"),
  clustering_distance_rows = "euclidean",
  row_names_gp = gpar(fontsize = 10, fontface = "italic"),
  column_names_gp = gpar(fontsize = 10),
  top_annotation = top_annot,
  cluster_columns = col_dend,
  row_title_rot = 0,

```

```

        row_dend_reorder = FALSE,
        row_title_gp = gpar(fontsize = 10, fontface = "bold"),
        show_parent_dend_line = FALSE,
        heatmap_legend_param = list(direction = "horizontal",
title_position = "topcenter",
                                at = c(-5, -2.5, 0, 2.5, 5),
                                labels = c("-5", "", "", "0", "",
"5"))))
draw(hm_cytokine, heatmap_legend_side = "right", annotation_legend_side = "bottom")

# T cell signalling transduction
hm_tcell <- Heatmap(mat_tcell, name = "Expressão",
                    col = col_fun,
                    column_title = "Transdução de sinal em linfócitos T",
                    column_title_gp = gpar(fontsize = 12, fontface = "bold"),
                    clustering_distance_rows = "euclidean",
                    row_names_gp = gpar(fontsize = 10, fontface = "italic"),
                    column_names_gp = gpar(fontsize = 10),
                    top_annotation = top_annot,
                    cluster_columns = col_dend,
                    row_title_rot = 0,
                    row_dend_reorder = FALSE,
                    row_title_gp = gpar(fontsize = 10, fontface = "bold"),
                    show_parent_dend_line = FALSE,
                    heatmap_legend_param = list(direction = "horizontal",
title_position = "topcenter",
                                at = c(-5, -2.5, 0, 2.5, 5),
                                labels = c("-5", "", "", "0", "",
", "5"))))
draw(hm_tcell, heatmap_legend_side = "right", annotation_legend_side = "bottom")

## ----- Dispersion and Side-by-side plots of each category -----
# Making new matrix
counts_vst.int.plot <- as.data.frame(assay(vst_intDEG))

# Checking rownames
identical(rownames(counts_vst.int.plot), as.vector(genes_int_deg$hgnc_symbol))#
FALSE

# Ordering datasets alphabetically
counts_vst.int.plot <- counts_vst.int.plot[order(rownames(counts_vst.int.plot)),]
genes_int_deg <- genes_int_deg[order(genes_int_deg$hgnc_symbol),]

identical(rownames(counts_vst.int.plot), as.vector(genes_int_deg$hgnc_symbol)) #
TRUE

# Adjusting rownames to match known gene names
rownames(counts_vst.int.plot) <- genes_int_deg$names_plot

# Adding category column to separate datasets
counts_vst.int.plot$categoria <- genes_int_deg$categoria

## Separating each category, annotating samples and converting to long format for
plotting

# Signalling
counts_vst_signalling <-
counts_vst.int.plot[counts_vst.int.plot$categoria=="Sinalização",1:28]
counts_vst_signalling <- as.data.frame(t(counts_vst_signalling))
counts_vst_signalling$Source <- pheno_data$source
counts_vst_signalling$Source <- str_replace_all(counts_vst_signalling$Source,
"healthy_skin", "CS")
counts_vst_signalling$Source <- str_replace_all(counts_vst_signalling$Source,
"CL_lesion", "LCL")
counts_vst_signalling <- melt(counts_vst_signalling, id.vars=c("source"),
value.name = "Expression")
colnames(counts_vst_signalling) <- c("Grupo", "Gene", "Expressao")
counts_vst_signalling$Grupo <- factor(counts_vst_signalling$Grupo, c("CS", "LCL"))
counts_vst_signalling$Grupo2 <- counts_vst_signalling$Grupo
counts_vst_signalling$Grupo2 <- str_replace_all(counts_vst_signalling$Grupo2, "CS",
"CS2")
counts_vst_signalling$Grupo2 <- str_replace_all(counts_vst_signalling$Grupo2,
"LCL", "LCL2")

```

```

counts_vst_signalling$Grupo2 <- factor(counts_vst_signalling$Grupo2, c("CS2",
"LCL2"))

# Proliferation and cell cycle control
counts_vst_prolif <- counts_vst.int.plot[counts_vst.int.plot$categoria=="Controle
de proliferaçao e ciclo celular",1:28]
counts_vst_prolif <- as.data.frame(t(counts_vst_prolif))
counts_vst_prolif$Source <- pheno_data$source
counts_vst_prolif$Source <- str_replace_all(counts_vst_prolif$Source,
"healthy_skin", "CS")
counts_vst_prolif$Source <- str_replace_all(counts_vst_prolif$Source, "CL_lesion",
"LCL")
counts_vst_prolif <- melt(counts_vst_prolif, id.vars=c("source"), value.name =
"Expression")
colnames(counts_vst_prolif) <- c("Grupo", "Gene", "Expressao")
counts_vst_prolif$Grupo <- factor(counts_vst_prolif$Grupo, c("CS", "LCL"))
counts_vst_prolif$Grupo2 <- counts_vst_prolif$Grupo
counts_vst_prolif$Grupo2 <- str_replace_all(counts_vst_prolif$Grupo2, "CS", "CS2")
counts_vst_prolif$Grupo2 <- str_replace_all(counts_vst_prolif$Grupo2, "LCL",
"LCL2")
counts_vst_prolif$Grupo2 <- factor(counts_vst_prolif$Grupo2, c("CS2", "LCL2"))

# Transcription factores
counts_vst_transc <- counts_vst.int.plot[counts_vst.int.plot$categoria=="Fatores de
Transcriçao",1:28]
counts_vst_transc <- as.data.frame(t(counts_vst_transc))
counts_vst_transc$Source <- pheno_data$source
counts_vst_transc$Source <- str_replace_all(counts_vst_transc$Source,
"healthy_skin", "CS")
counts_vst_transc$Source <- str_replace_all(counts_vst_transc$Source, "CL_lesion",
"LCL")
counts_vst_transc <- melt(counts_vst_transc, id.vars=c("source"), value.name =
"Expression")
colnames(counts_vst_transc) <- c("Grupo", "Gene", "Expressao")
counts_vst_transc$Grupo <- factor(counts_vst_transc$Grupo, c("CS", "LCL"))
counts_vst_transc$Grupo2 <- counts_vst_transc$Grupo
counts_vst_transc$Grupo2 <- str_replace_all(counts_vst_transc$Grupo2, "CS", "CS2")
counts_vst_transc$Grupo2 <- str_replace_all(counts_vst_transc$Grupo2, "LCL",
"LCL2")
counts_vst_transc$Grupo2 <- factor(counts_vst_transc$Grupo2, c("CS2", "LCL2"))

# Cytotoxicity
counts_vst_citotox <-
counts_vst.int.plot[counts_vst.int.plot$categoria=="Citotoxicidade",1:28]
counts_vst_citotox <- as.data.frame(t(counts_vst_citotox))
counts_vst_citotox$Source <- pheno_data$source
counts_vst_citotox$Source <- str_replace_all(counts_vst_citotox$Source,
"healthy_skin", "CS")
counts_vst_citotox$Source <- str_replace_all(counts_vst_citotox$Source,
"CL_lesion", "LCL")
counts_vst_citotox <- melt(counts_vst_citotox, id.vars=c("source"), value.name =
"Expression")
colnames(counts_vst_citotox) <- c("Grupo", "Gene", "Expressao")
counts_vst_citotox$Grupo <- factor(counts_vst_citotox$Grupo, c("CS", "LCL"))
counts_vst_citotox$Grupo2 <- counts_vst_citotox$Grupo
counts_vst_citotox$Grupo2 <- str_replace_all(counts_vst_citotox$Grupo2, "CS",
"CS2")
counts_vst_citotox$Grupo2 <- str_replace_all(counts_vst_citotox$Grupo2, "LCL",
"LCL2")
counts_vst_citotox$Grupo2 <- factor(counts_vst_citotox$Grupo2, c("CS2", "LCL2"))

# Costimulatory and homing receptors
counts_vst_coest <- counts_vst.int.plot[counts_vst.int.plot$categoria=="Receptores
coestimulatórios e de homing",1:28]
counts_vst_coest <- as.data.frame(t(counts_vst_coest))
counts_vst_coest$Source <- pheno_data$source
counts_vst_coest$Source <- str_replace_all(counts_vst_coest$Source, "healthy_skin",
"CS")
counts_vst_coest$Source <- str_replace_all(counts_vst_coest$Source, "CL_lesion",
"LCL")
counts_vst_coest <- melt(counts_vst_coest, id.vars=c("source"), value.name =
"Expression")
colnames(counts_vst_coest) <- c("Grupo", "Gene", "Expressao")
counts_vst_coest$Grupo <- factor(counts_vst_coest$Grupo, c("CS", "LCL"))
counts_vst_coest$Grupo2 <- counts_vst_coest$Grupo
counts_vst_coest$Grupo2 <- str_replace_all(counts_vst_coest$Grupo2, "CS", "CS2")

```

```
counts_vst_coest$Grupo2 <- str_replace_all(counts_vst_coest$Grupo2, "LCL", "LCL2")
counts_vst_coest$Grupo2 <- factor(counts_vst_coest$Grupo2, c("CS2", "LCL2"))
```

Innate signalling molecules

```
counts_vst_innate <- counts_vst.int.plot[counts_vst.int.plot$categoria=="Moléculas
de sinalização inata",1:28]
counts_vst_innate <- as.data.frame(t(counts_vst_innate))
counts_vst_innate$Source <- pheno_data$source
counts_vst_innate$Source <- str_replace_all(counts_vst_innate$Source,
"healthy_skin", "CS")
counts_vst_innate$Source <- str_replace_all(counts_vst_innate$Source, "CL_lesion",
"LCL")
counts_vst_innate <- melt(counts_vst_innate, id.vars=c("Source"), value.name =
"Expression")
colnames(counts_vst_innate) <- c("Grupo", "Gene", "Expressao")
counts_vst_innate$Grupo <- factor(counts_vst_innate$Grupo, c("CS", "LCL"))
counts_vst_innate$Grupo2 <- counts_vst_innate$Grupo
counts_vst_innate$Grupo2 <- str_replace_all(counts_vst_innate$Grupo2, "CS", "CS2")
counts_vst_innate$Grupo2 <- str_replace_all(counts_vst_innate$Grupo2, "LCL",
"LCL2")
counts_vst_innate$Grupo2 <- factor(counts_vst_innate$Grupo2, c("CS2", "LCL2"))
```

Cytokine receptors

```
counts_vst_cytokine <-
counts_vst.int.plot[counts_vst.int.plot$categoria=="Receptores de citocinas",1:28]
counts_vst_cytokine <- as.data.frame(t(counts_vst_cytokine))
counts_vst_cytokine$Source <- pheno_data$source
counts_vst_cytokine$Source <- str_replace_all(counts_vst_cytokine$Source,
"healthy_skin", "CS")
counts_vst_cytokine$Source <- str_replace_all(counts_vst_cytokine$Source,
"CL_lesion", "LCL")
counts_vst_cytokine <- melt(counts_vst_cytokine, id.vars=c("Source"), value.name =
"Expression")
colnames(counts_vst_cytokine) <- c("Grupo", "Gene", "Expressao")
counts_vst_cytokine$Grupo <- factor(counts_vst_cytokine$Grupo, c("CS", "LCL"))
counts_vst_cytokine$Grupo2 <- counts_vst_cytokine$Grupo
counts_vst_cytokine$Grupo2 <- str_replace_all(counts_vst_cytokine$Grupo2, "CS",
"CS2")
counts_vst_cytokine$Grupo2 <- str_replace_all(counts_vst_cytokine$Grupo2, "LCL",
"LCL2")
counts_vst_cytokine$Grupo2 <- factor(counts_vst_cytokine$Grupo2, c("CS2", "LCL2"))
```

T cell signalling transduction

```
counts_vst_tcell <- counts_vst.int.plot[counts_vst.int.plot$categoria=="Transdução
de sinal em linfócitos T",1:28]
counts_vst_tcell <- as.data.frame(t(counts_vst_tcell))
counts_vst_tcell$Source <- pheno_data$source
counts_vst_tcell$Source <- str_replace_all(counts_vst_tcell$Source, "healthy_skin",
"CS")
counts_vst_tcell$Source <- str_replace_all(counts_vst_tcell$Source, "CL_lesion",
"LCL")
counts_vst_tcell <- melt(counts_vst_tcell, id.vars=c("Source"), value.name =
"Expression")
colnames(counts_vst_tcell) <- c("Grupo", "Gene", "Expressao")
counts_vst_tcell$Grupo <- factor(counts_vst_tcell$Grupo, c("CS", "LCL"))
counts_vst_tcell$Grupo2 <- counts_vst_tcell$Grupo
counts_vst_tcell$Grupo2 <- str_replace_all(counts_vst_tcell$Grupo2, "CS", "CS2")
counts_vst_tcell$Grupo2 <- str_replace_all(counts_vst_tcell$Grupo2, "LCL", "LCL2")
counts_vst_tcell$Grupo2 <- factor(counts_vst_tcell$Grupo2, c("CS2", "LCL2"))
```

NK ligands

```
counts_vst_NKligands <-
counts_vst.int.plot[counts_vst.int.plot$categoria=="Ligantes de células NK",1:28]
counts_vst_NKligands <- as.data.frame(t(counts_vst_NKligands))
counts_vst_NKligands$Source <- pheno_data$source
counts_vst_NKligands$Source <- str_replace_all(counts_vst_NKligands$Source,
"healthy_skin", "CS")
counts_vst_NKligands$Source <- str_replace_all(counts_vst_NKligands$Source,
"CL_lesion", "LCL")
counts_vst_NKligands <- melt(counts_vst_NKligands, id.vars=c("Source"), value.name =
"Expression")
colnames(counts_vst_NKligands) <- c("Grupo", "Gene", "Expressao")
counts_vst_NKligands$Grupo <- factor(counts_vst_NKligands$Grupo, c("CS", "LCL"))
counts_vst_NKligands$Grupo2 <- counts_vst_NKligands$Grupo
counts_vst_NKligands$Grupo2 <- str_replace_all(counts_vst_NKligands$Grupo2, "CS",
"CS2")
```

```

counts_vst_NKligands$Grupo2 <- str_replace_all(counts_vst_NKligands$Grupo2, "LCL",
"LCL2")
counts_vst_NKligands$Grupo2 <- factor(counts_vst_NKligands$Grupo2, c("CS2",
"LCL2"))

# NK receptors
counts_vst_NKreceptors <-
counts_vst.int.plot[counts_vst.int.plot$categoria=="Receptores de células NK",1:28]
counts_vst_NKreceptors <- as.data.frame(t(counts_vst_NKreceptors))
counts_vst_NKreceptors$Source <- pheno_data$source
counts_vst_NKreceptors$Source <- str_replace_all(counts_vst_NKreceptors$Source,
"healthy_skin", "CS")
counts_vst_NKreceptors$Source <- str_replace_all(counts_vst_NKreceptors$Source,
"CL_lesion", "LCL")
counts_vst_NKreceptors <- melt(counts_vst_NKreceptors, id.vars=c("Source"),
value.name = "Expression")
colnames(counts_vst_NKreceptors) <- c("Grupo", "Gene", "Expressao")
counts_vst_NKreceptors$Grupo <- factor(counts_vst_NKreceptors$Grupo, c("CS",
"LCL"))
counts_vst_NKreceptors$Grupo2 <- counts_vst_NKreceptors$Grupo
counts_vst_NKreceptors$Grupo2 <- str_replace_all(counts_vst_NKreceptors$Grupo2,
"CS", "CS2")
counts_vst_NKreceptors$Grupo2 <- str_replace_all(counts_vst_NKreceptors$Grupo2,
"LCL", "LCL2")
counts_vst_NKreceptors$Grupo2 <- factor(counts_vst_NKreceptors$Grupo2, c("CS2",
"LCL2"))

# SASP
counts_vst_SASP <- counts_vst.int.plot[counts_vst.int.plot$categoria=="SASP",1:28]
counts_vst_SASP <- as.data.frame(t(counts_vst_SASP))
counts_vst_SASP$Source <- pheno_data$source
counts_vst_SASP$Source <- str_replace_all(counts_vst_SASP$Source, "healthy_skin",
"CS")
counts_vst_SASP$Source <- str_replace_all(counts_vst_SASP$Source, "CL_lesion",
"LCL")
counts_vst_SASP <- melt(counts_vst_SASP, id.vars=c("Source"), value.name =
"Expression")
colnames(counts_vst_SASP) <- c("Grupo", "Gene", "Expressao")
counts_vst_SASP$Grupo <- factor(counts_vst_SASP$Grupo, c("CS", "LCL"))
counts_vst_SASP$Grupo2 <- counts_vst_SASP$Grupo
counts_vst_SASP$Grupo2 <- str_replace_all(counts_vst_SASP$Grupo2, "CS", "CS2")
counts_vst_SASP$Grupo2 <- str_replace_all(counts_vst_SASP$Grupo2, "LCL", "LCL2")
counts_vst_SASP$Grupo2 <- factor(counts_vst_SASP$Grupo2, c("CS2", "LCL2"))

## Plotting side-by-side dispersion plots and saving in high quality format

# Signalling
counts_vst_signalling$Gene <- str_replace_all(counts_vst_signalling$Gene, " \\(",
"\\n\\(")

tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_Sigalling.tiff",
units="in", width=9, height=5, res=300)

ggplot(data=counts_vst_signalling, aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 0, face = "italic", size = 9),
axis.title.x=element_blank(),
axis.title.y = element_text(size = 12), plot.title = element_text(size =
12, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "sinalização", x = "", y = "Expressão normalizada em log2 (vst)")

dev.off()

# Proliferation and cell cycle control
counts_vst_prolif$Gene <- str_replace_all(counts_vst_prolif$Gene, "cyclin",
"Ciclina")
counts_vst_prolif$Gene <- str_replace_all(counts_vst_prolif$Gene, " \\(", " \\n\\(")

```

```

tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_Proliferation.tiff",
units="in", width=9, height=5, res=300)

ggplot(data=counts_vst_prolif, aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 0, face = "italic", size = 9),
axis.title.x=element_blank(),
axis.title.y = element_text(size = 12), plot.title = element_text(size =
12, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "Controle de proliferação e ciclo celular", x = "", y = "Expressão
normalizada em log2 (vst)")

dev.off()

# Transcription factors
counts_vst_transc$Gene <- str_replace_all(counts_vst_transc$Gene, " \\(", "\\n\\(")

tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_TranscFactors.tiff",
units="in", width=9, height=5, res=300)

ggplot(data=counts_vst_transc, aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 0, face = "italic", size = 8),
axis.title.x=element_blank(),
axis.title.y = element_text(size = 12), plot.title = element_text(size =
12, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "Fatores de transcrição", x = "", y = "Expressão normalizada em log2
(vst)")

dev.off()

# Cytotoxicity
counts_vst_citotox$Gene <- str_replace_all(counts_vst_citotox$Gene, " \\(",
"\\n\\(")
counts_vst_citotox$Gene <- str_replace_all(counts_vst_citotox$Gene, "Gramzyme",
"Granzima")
counts_vst_citotox$Gene <- str_replace_all(counts_vst_citotox$Gene, "Perforin",
"Perforina")

tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_Cytotoxicity.tiff",
units="in", width=9, height=5, res=300)

ggplot(data=counts_vst_citotox, aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 0, face = "italic", size = 8),
axis.title.x=element_blank(),
axis.title.y = element_text(size = 12), plot.title = element_text(size =
12, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "Citotoxicidade", x = "", y = "Expressão normalizada em log2 (vst)")

dev.off()

# Coestimulatory and homing receptors

```

```

counts_vst_coest$Gene <- str_replace_all(counts_vst_coest$Gene, " \\(", "\\n\\(")
tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_Coestimulatory.tiff",
units="in", width=9, height=5, res=300)

ggplot(data=counts_vst_coest, aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 0, face = "italic", size = 8),
axis.title.x=element_blank(),
axis.title.y = element_text(size = 12), plot.title = element_text(size =
12, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "Moléculas coestimulatórias e de homing", x = "", y = "Expressão
normalizada em log2 (vst)")

dev.off()

# Innate signalling molecules
counts_vst_innate$Gene <- str_replace_all(counts_vst_innate$Gene, " \\(", "\\n\\(")
tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_Innate.tiff", units="in",
width=9, height=5, res=300)

ggplot(data=counts_vst_innate, aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 0, face = "italic", size = 8),
axis.title.x=element_blank(),
axis.title.y = element_text(size = 12), plot.title = element_text(size =
12, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "Moléculas de sinalização inata", x = "", y = "Expressão normalizada
em log2 (vst)")

dev.off()

# Cytokine receptors
counts_vst_cytokine$Gene <- str_replace_all(counts_vst_cytokine$Gene, " \\(",
"\\n\\(")
tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_Cytokines.tiff",
units="in", width=9, height=5, res=300)

ggplot(data=counts_vst_cytokine, aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 0, face = "italic", size = 8),
axis.title.x=element_blank(),
axis.title.y = element_text(size = 12), plot.title = element_text(size =
12, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "Receptores de citocinas", x = "", y = "Expressão normalizada em
log2 (vst)")

dev.off()

# T-cell signalling transduction
counts_vst_tcell$Gene <- str_replace_all(counts_vst_tcell$Gene, " \\(", "\\n\\(")

```

```

tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_Tcell.tiff", units="in",
width=9, height=5, res=300)

ggplot(data=counts_vst_tcell, aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 0, face = "italic", size = 8),
axis.title.x=element_blank(),
axis.title.y = element_text(size = 12), plot.title = element_text(size =
12, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "Transdução de sinal em linfócitos T", x = "", y = "Expressão
normalizada em log2 (vst)")

dev.off()

# NK ligands
counts_vst_NKligands$Gene <- str_replace_all(counts_vst_NKligands$Gene, " \\(",
"\n\\(")

tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_NKligands.tiff",
units="in", width=9, height=5, res=300)

ggplot(data=counts_vst_NKligands, aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 0, face = "italic", size = 8),
axis.title.x=element_blank(),
axis.title.y = element_text(size = 12), plot.title = element_text(size =
12, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "Ligantes de células NK", x = "", y = "Expressão normalizada em log2
(vst)")

dev.off()

# NK receptors 6x15
counts_vst_NKreceptors$Gene <- str_replace_all(counts_vst_NKreceptors$Gene, " \\(",
"\n\\(")

tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_NKreceptors.tiff",
units="in", width=15, height=8, res=300)

ggplot(data=counts_vst_NKreceptors, aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 0, face = "italic", size = 8),
axis.title.x=element_blank(),
axis.title.y = element_text(size = 12), plot.title = element_text(size =
12, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "Receptores de células NK", x = "", y = "Expressão normalizada em
log2 (vst)")

dev.off()

# SASP 6x15
counts_vst_SASP$Gene <- str_replace_all(counts_vst_SASP$Gene, " \\(", "\n\\(")

```

```

tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_SASP1.tiff", units="in",
width=15, height=7, res=300)

ggplot(data=counts_vst_SASP[1:504,], aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 90, face = "italic", size = 14),
axis.title.x=element_blank(),
axis.text.y = element_text(angle = 0, face = "italic", size = 14),
axis.title.y = element_text(size = 16), plot.title = element_text(size =
16, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "SASP", x = "", y = "Expressão normalizada em log2 (vst)")

dev.off()

tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_SASP2.tiff", units="in",
width=15, height=7, res=300)

ggplot(data=counts_vst_SASP[505:1008,], aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 90, face = "italic", size = 14),
axis.title.x=element_blank(),
axis.text.y = element_text(angle = 0, face = "italic", size = 14),
axis.title.y = element_text(size = 16), plot.title = element_text(size =
16, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "SASP", x = "", y = "Expressão normalizada em log2 (vst)")

dev.off()

tiff("results/dissertacao_att/individual_sbs/Dotplot_VST_SASP3.tiff", units="in",
width=15, height=7, res=300)

ggplot(data=counts_vst_SASP[1009:1456,], aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 90, face = "italic", size = 14),
axis.title.x=element_blank(),
axis.text.y = element_text(angle = 0, face = "italic", size = 14),
axis.title.y = element_text(size = 16), plot.title = element_text(size =
16, face = "bold", hjust = 0.5)) +
  scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
  labs(title = "SASP", x = "", y = "Expressão normalizada em log2 (vst)")

dev.off()

ggplot(data=counts_vst_SASP, aes(x=Gene, y=Expressao, color=Grupo)) +
  stat_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), geom='errorbar', width =
0.4, position = position_dodge(width = 0.75), show.legend = FALSE) +
  geom_boxplot(aes(x=Gene, y=Expressao, color=Grupo2), show.legend = FALSE,
outlier.shape = NA) +
  geom_point(position=position_jitterdodge(jitter.width = 0.4, jitter.height = 0),
size = 3) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 90, face = "italic", size = 14),
axis.title.x=element_blank(),

```

```

axis.text.y = element_text(angle = 0, face = "italic", size = 14),
axis.title.y = element_text(size = 16), plot.title = element_text(size =
16, face = "bold", hjust = 0.5)) +
scale_colour_manual(breaks = c("CS", "LCL"), values = c("CS" = "#33C4FF", "LCL" =
"#FD4646", "CS2" = "black", "LCL2" = "black")) +
labs(title = "", x = "", y = "Expressão normalizada em log2 (vst)")

##----- Saving high quality individual groups heatmaps -----

tiff("results/dissertacao_att/individual_hm_tiff/Heatmap_VST_TCell.tiff",
units="in", width=7, height=5, res=300)
draw(hm_tcell, heatmap_legend_side = "right", annotation_legend_side = "bottom")
dev.off()

tiff("results/dissertacao_att/individual_hm_tiff/Heatmap_VST_Cytokine.tiff",
units="in", width=7, height=5, res=300)
draw(hm_cytokine, heatmap_legend_side = "right", annotation_legend_side = "bottom")
dev.off()

tiff("results/dissertacao_att/individual_hm_tiff/Heatmap_VST_Innate.tiff",
units="in", width=7, height=5, res=300)
draw(hm_innate, heatmap_legend_side = "right", annotation_legend_side = "bottom")
dev.off()

tiff("results/dissertacao_att/individual_hm_tiff/Heatmap_VST_Coest.tiff",
units="in", width=7, height=5, res=300)
draw(hm_coest, heatmap_legend_side = "right", annotation_legend_side = "bottom")
dev.off()

tiff("results/dissertacao_att/individual_hm_tiff/Heatmap_VST_Cytotoxicity.tiff",
units="in", width=7, height=5, res=300)
draw(hm_citotox, heatmap_legend_side = "right", annotation_legend_side = "bottom")
dev.off()

tiff("results/dissertacao_att/individual_hm_tiff/Heatmap_VST_TranscFactors.tiff",
units="in", width=7, height=5, res=300)
draw(hm_transc, heatmap_legend_side = "right", annotation_legend_side = "bottom")
dev.off()

tiff("results/dissertacao_att/individual_hm_tiff/Heatmap_VST_Proliferation.tiff",
units="in", width=7, height=5, res=300)
draw(hm_prolif, heatmap_legend_side = "right", annotation_legend_side = "bottom")
dev.off()

tiff("results/dissertacao_att/individual_hm_tiff/Heatmap_VST_Signalling.tiff",
units="in", width=7, height=5, res=300)
draw(hm_signalling, heatmap_legend_side = "right", annotation_legend_side =
"bottom")
dev.off()

tiff("results/dissertacao_att/individual_hm_tiff/Heatmap_VST_SASP.tiff",
units="in", width=7, height=9, res=300)
draw(hm_SASP, heatmap_legend_side = "right", annotation_legend_side = "bottom")
dev.off()

tiff("results/dissertacao_att/individual_hm_tiff/Heatmap_VST_NKreceptors.tiff",
units="in", width=7, height=5, res=300)
draw(hm_NKreceptors, heatmap_legend_side = "right", annotation_legend_side =
"bottom")
dev.off()

tiff("results/dissertacao_att/individual_hm_tiff/Heatmap_VST_NKligands.tiff",
units="in", width=7, height=5, res=300)
draw(hm_NKligands, heatmap_legend_side = "right", annotation_legend_side =
"bottom")
dev.off()

##----- Plotting custom PCA -----

# Extracting data from pca function
pcaData <- plotPCA(vst, intgroup = "source", returnData = TRUE)
percentVar <- round(100 * attr(pcaData, "percentVar"))

# Plotting and saving in high quality
tiff("results/dissertacao_att/hqplots/PCA_VST_PT.tiff", units="in", width=6,
height=3, res=300)

```

```

ggplot(pcaData, aes(PC1, PC2, colour = group)) +
  geom_point(size = 4) +
  scale_fill_manual(values = c("#33C4FF", "#FD4646")) +
  labs(colour = "Source") +
  scale_colour_manual(labels = c("CS", "LCL"), name = "Grupo",
                      values = c("#33C4FF", "#FD4646"),
                      guide = guide_legend(reverse = TRUE)) +
  xlab(paste0("PC1: ", percentVar[1], "% variância")) +
  ylab(paste0("PC2: ", percentVar[2], "% variância")) +
  coord_fixed()

dev.off()

##----- Saving results from DESeq analysis -----

res_signif.int <- subset(res_signif, rownames(res_signif) %in%
genes_int$hgnc_symbol[genes_int$is_DE==TRUE])

identical(rownames(res_signif.int), sort(genes_list_true))

res_signif.int <- as.data.frame(res_signif.int)
res_signif.int$fold_change <- ifelse(res_signif.int$log2FoldChange > 0,
2^res_signif.int$log2FoldChange, -1 / (2 ^res_signif.int$log2FoldChange))

res_total <- as.data.frame(res_total_int)

# Converting log2 fold change to real fold change
res_total$fold_change <- ifelse(res_total$log2FoldChange > 0,
2^res_total$log2FoldChange, -1 / (2 ^res_total$log2FoldChange))

# Writing files
write.xlsx(res_signif.int, "results/Genes_of_Interest_LFC.xlsx")
write.xlsx(res_total, "results/res_total.xlsx")

##----- Gene Ontology using clusterProfiler -----

## Defining geneLists

# All significant DEGs
all_geneList <- as.numeric(res_signif$log2FoldChange)
names(all_geneList) <- as.character(rownames(res_signif))
all_geneList <- sort(all_geneList, decreasing = TRUE)

# Genes of Interest
int_geneList <- as.numeric(res_signif.int$log2FoldChange)
names(int_geneList) <- as.character(rownames(res_signif.int))
int_geneList <- sort(int_geneList, decreasing = TRUE)

head(all_geneList)
head(int_geneList)

## Gene Ontology (GO) using Over-representation test

# Biological Processes - All DEGs
ego_BP_all <- enrichGO(gene = names(all_geneList), orgDb = org.Hs.eg.db, keyType =
"SYMBOL",
ont = "BP", pAdjustMethod = "BH", pvalueCutoff = 0.05,
qvalueCutoff = 0.05,
universe = rownames(dds))

head(ego_BP_all)
dim(ego_BP_all)

ego_BP_all_simp <- simplify(ego_BP_all, cutoff = 0.7, by = "p.adjust", select_fun =
min)

head(ego_BP_all_simp)
dim(ego_BP_all_simp)

# write.xlsx(ego_BP_all, "results/gene_ontology/OR_Enrichment_BP_All-DEGs.xlsx",
row.names = FALSE, sheetName = "BP All DEGs")

## ----- Gene Ontology Plots -----
---
```

```

# Adjusting name for plot
ego_BP_all@result$Description[3] <- "adaptive immune response based on somatic
recombination of \immune receptors built from immunoglobulin superfamily domains"

# Barplot
barplot(ego_BP_all, showCategory = 30, font.size = 8)

# Dotplot
tiff("results/dissertacao_att/hqplots/Dotplot_OR_30.tiff", units="in", width=6,
height=7.5, res=300)

dotplot(ego_BP_all, showCategory = 30, x = "Count", font.size = 8)

dev.off()

##----- volcano plots using EnhancedVolcano -----

head(res)
summary(res)

# Test
EnhancedVolcano(res, x = "log2FoldChange", y = "padj", lab = rownames(res), title =
"Expressão gênica diferencial (DESeq2)",
                subtitle = NULL, pCutoff = 0.05, FCcutoff = 0)

# Custom coloring
res_custom <- as.data.frame(res)
res_custom$keyvals <- ""
res_custom$keyvals[res_custom$padj < 0.05 & res_custom$log2FoldChange > 0] <- "red"
res_custom$keyvals[res_custom$padj < 0.05 & res_custom$log2FoldChange < 0] <-
"blue"
res_custom$keyvals[is.na(res_custom$padj)] <- "grey30"
res_custom$keyvals[res_custom$keyvals==""] <- "grey30"

keyvals <- res_custom$keyvals

names(keyvals)[keyvals == "grey30"] <- "Não significativo"
names(keyvals)[keyvals == "blue"] <- "Regulação negativa"
names(keyvals)[keyvals == "red"] <- "Regulação positiva"

EnhancedVolcano(res, x = "log2FoldChange", y = "padj", lab = rownames(res), title =
"Expressão gênica diferencial (DESeq2)",
                subtitle = NULL, pCutoff = 0.05, FCcutoff = 1, colCustom = keyvals,
caption = NULL, legendPosition = "right",
                legendLabSize = 12, titleLabSize = 14, selectLab =
genes_int_deg$hgnc_symbol, labFace = "italic",
                drawConnectors = TRUE, lengthConnectors = unit(0.0001, "npc")) +
  theme(axis.text.x = element_text(size=14), axis.text.y = element_text(size=14),
        axis.title = element_text(size=14,face="bold"))

## Only significant genes
res_custom_signif <- as.data.frame(res_signif)
res_custom_signif$keyvals <- ""
res_custom_signif$keyvals[res_custom_signif$padj < 0.05 &
res_custom_signif$log2FoldChange > 0] <- "red"
res_custom_signif$keyvals[res_custom_signif$padj < 0.05 &
res_custom_signif$log2FoldChange < 0] <- "blue"

keyvals_signif <- res_custom_signif$keyvals
names(keyvals_signif)[keyvals_signif == "blue"] <- "Regulação negativa"
names(keyvals_signif)[keyvals_signif == "red"] <- "Regulação positiva"

# Saving high quality plot
tiff("results/dissertacao_att/hqplots/DEGs_volcanoPlot.tiff", units="in", width=8,
height=5, res=300)

EnhancedVolcano(res_signif, x = "log2FoldChange", y = "padj", lab =
rownames(res_signif), title = "Expressão gênica diferencial (DESeq2)",
                subtitle = NULL, pCutoff = 0.05, FCcutoff = 1, colCustom =
keyvals_signif, caption = NULL, legendPosition = "right",
                legendLabSize = 12, titleLabSize = 14, selectLab =
genes_int_deg$hgnc_symbol, labFace = "italic") +
  theme(axis.text.x = element_text(size=14), axis.text.y = element_text(size=14),
        axis.title = element_text(size=14,face="bold"), legend.text =
element_text(size = 8))

```

```

dev.off()

##----- ImmQuant deconvolution analysis -----

## Importing scores from software
vst_immquant_scores <- as.data.frame(read_delim("ImmQuant/vst_immquant_scores.txt",
"\t", escape_double = FALSE, trim_ws = TRUE))
rownames(vst_immquant_scores) <- vst_immquant_scores[,1]
vst_immquant_scores <- vst_immquant_scores[,-1]

## Plotting

# Transposing
scores_plot <- vst_immquant_scores
rownames(scores_plot) <- scores_plot[,1]
scores_plot <- scores_plot[,-1]
scores_plot <- as.data.frame(t(scores_plot))

# Annotating source
scores_plot$Source <- ""
scores_plot$Source[1:7] <- "Healthy control"
scores_plot$Source[8:28] <- "CL lesion"

# Converting to long format
scores_plot_long <- melt(scores_plot, id.vars=c("Source"), value.name = "Score")
colnames(scores_plot_long) <- c("Source", "Cell", "Score")

# Plotting
ggplot(scores_plot_long, aes(x = Source, fill = Source, colour = Source, y =
Score)) +
  geom_dotplot(binaxis = "y", stackdir = "center") +
  labs(title = "ImmQuant deconvolution of immune cells", x = "", y = "ImmQuant
Score") +
  theme(plot.title = element_text(hjust = 0.5), axis.text.x = element_blank(),
        axis.ticks.x = element_blank(), strip.text.x = element_text(size = 8)) +
  scale_x_discrete(limits = c("Healthy control", "CL lesion")) +
  scale_fill_manual(values = c("#FD4646", "#33C4FF")) +
  scale_colour_manual(values = c("#FD4646", "#33C4FF")) +
  facet_wrap(~ Cell)

## Plotting using ComplexHeatmaps

immquant_heatmap <- data.matrix(vst_immquant_scores[,2:29])
rownames(immquant_heatmap) <- vst_immquant_scores[,1]

col_immquant <- colorRamp2(c(-1, 0, 1), c("blue", "white", "red"))

hm_immquant <- Heatmap(immquant_heatmap, name = "Pontuação relativa",
  column_title = "Score de deconvolução celular (ImmQuant)",
  col = col_immquant,
  clustering_distance_rows = "euclidean",
  column_title_gp = gpar(fontsize = 12, fontface = "bold"),
  row_names_gp = gpar(fontsize = 9),
  column_names_gp = gpar(fontsize = 9),
  cluster_columns = FALSE,
  bottom_annotation = top_annot,
  row_title_rot = 0,
  show_row_dend = FALSE,
  row_title_gp = gpar(fontsize = 10, fontface = "bold"),
  heatmap_legend_param = list(direction = "horizontal",
title_position = "topcenter",
                                at = c(-1, -0.5, 0, 0.5, 1),
                                labels = c("-1", "", "0", "" ,
"1"))))

## Saving high quality heatmap
tiff("results/dissertacao_att/hqplots/ImmQuantRelativeScores.tiff", units="in",
width=10, height=8, res=300)

draw(hm_immquant, heatmap_legend_side = "bottom", annotation_legend_side =
"bottom")

dev.off()

## ----- Gene x Cell correlation -----

## Testing a few combinations

```

```

# Defining genes and cells of interest
corr_genes <- c("GZMB", "PRF1", "KLRK1", "SESN2", "FASLG", "FAS", "IL6", "CXCL8")
corr_cells <- c("CD4+ Effector Memory", "CD8+ Effector Memory", "CD8+ Effector
Memory RA", "Mature NK cell_CD56- CD16- CD3-",
               "Mature NK cell_CD56- CD16+ CD3-", "Mature NK cell_CD56+ CD16+ CD3-
", "Granulocyte (Neutrophil)", "Monocyte")

# VST assay
vst_intDEG_assay <- as.data.frame(assay(vst))

# Joining genes + cell scores
cells_and_genes <- rbind(vst_immquant_scores[corr_cells,],
vst_intDEG_assay[corr_genes,])

# Transposing
cells_and_genes <- t(cells_and_genes)

# Creating correlation matrix
cormat <- round(cor(cells_and_genes, method = "spearman"), 2)

# p-values matrix
p.mat <- cor_pmat(cells_and_genes)

ggcorrplot(cormat, hc.order = FALSE, outline.col = "white", p.mat = p.mat,
sig.level = 0.05, insig = "pch", type = "lower")

## ----- Actual correlation analysis -----

# Importing genes and cell types to correlate
corr_genes <- read_excel("checking_genes/genes-cells_to_correlate.xlsx", sheet =
"Genes")
corr_cells <- vst_immquant_scores[c(2,3,4,5,6,26,27,28), 1]

## Matching gene aliases

# Subsetting genes of interest table
genes_to_corr <- genes_int_deg[genes_int_deg$hgnc_symbol %in% corr_genes$`Gene
Symbol`,]
genes_to_corr <- rbind(genes_to_corr, c("IFN-??", "IFNG", "Cytokine", "",
"Citocina", TRUE, TRUE, "IFNG (IFN-??)"))
genes_to_corr <- rbind(genes_to_corr, c("TNF-??", "TNF", "Cytokine", "",
"Citocina", TRUE, TRUE, "TNF (TNF-??)"))
genes_to_corr <- rbind(genes_to_corr, c("hTERT", "TERT", "Telomerase", "",
"Telomerase", TRUE, TRUE, "TERT (hTERT)"))

# Checking if the datasets match
corr_genes <- corr_genes[order(corr_genes$`Gene Symbol`),]
genes_to_corr <- genes_to_corr[order(genes_to_corr$hgnc_symbol),]

identical(corr_genes$`Gene Symbol`, genes_to_corr$hgnc_symbol) # TRUE

# Adding additional data to corr_genes
corr_genes$name <- genes_to_corr$name
corr_genes$name_plot <- genes_to_corr$names_plot

# Removing dataset that will not be needed
rm(genes_to_corr)

# Joining gene expression + cell scores
corr_cells <- vst_immquant_scores[vst_immquant_scores$X1 %in% corr_cells,]
rownames(corr_cells) <- corr_cells$X1
corr_cells <- corr_cells[,-1]

corr_genes_bind <- vst_intDEG_assay[corr_genes$`Gene Symbol`,]
rownames(corr_genes_bind) <- corr_genes$name_plot

cells_and_genes <- rbind(corr_cells, corr_genes_bind)

# Transposing
cells_and_genes <- t(cells_and_genes)

# Creating correlation matrix
cormat <- round(cor(cells_and_genes, method = "spearman"), 2)

# p-values matrix

```

```

p.mat <- cor_pmat(cells_and_genes)

# ggcorrplot(cormat, hc.order = FALSE, outline.col = "white", p.mat = p.mat,
sig.level = 0.05, insig = "pch", type = "lower",
# pch = 4, pch.cex = 2, tl.cex = 10, tl.srt = 90)

### Using ggplot2

## Functions

# Get lower triangle of the correlation matrix
get_lower_tri<-function(cormat){
  cormat[upper.tri(cormat)] <- NA
  return(cormat)
}
# Get upper triangle of the correlation matrix
get_upper_tri <- function(cormat){
  cormat[lower.tri(cormat)]<- NA
  return(cormat)
}

# Getting upper triangle of matrices
upper_cormat <- get_upper_tri(cormat)
upper_pmat <- get_upper_tri(p.mat)

# Melting matrices
melted_cormat <- melt(upper_cormat, na.rm = TRUE)
melted_pmat <- melt(upper_pmat, na.rm = TRUE)

# Coding p-values
melted_pmat$pcode <- ""
for (i in 1:length(melted_pmat$value)) {
  if (melted_pmat[i,3] < 0.05 & melted_pmat[i,3] >= 0.01) {
    melted_pmat[i,4] <- "***"
  } else if (melted_pmat[i,3] < 0.01 & melted_pmat[i,3] >= 0.001) {
    melted_pmat[i,4] <- "**"
  } else if (melted_pmat[i,3] < 0.001 & melted_pmat[i,3] > 0) {
    melted_pmat[i,4] <- "*"
  }
}

melted_cormat$pcode <- melted_pmat$pcode

## Plotting and saving high quality image

tiff("results/dissertacao_att/hqplots/CellxGene_CorMap.tiff", units="in", width=10,
height=10, res=300)

ggplot(data = melted_cormat, aes(Var2, Var1, fill = value))+
  geom_tile(color = "white") +
  labs(title = "Mapa de Correlação\n Genes x Populações Celulares", x = "", y = "")
+
  scale_fill_gradient2(low = "blue", high = "red", mid = "white",
midpoint = 0, limit = c(-1,1), space = "Lab",
name="Correlação\nde Spearman") +
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 90, vjust = 0.3,
size = 10, hjust = 1))+
  coord_fixed() +
  geom_text(aes(Var2, Var1, label = pcode), color = "black", size = 3)+
  theme(axis.title.x = element_blank(), axis.title.y = element_blank(),
panel.grid.major = element_blank(), panel.border = element_blank(),
panel.background = element_blank(), axis.ticks = element_blank(),
legend.justification = c(0.5, 0), legend.position = c(0.3, 0.8),
legend.direction = "horizontal", legend.text=element_text(size=8),
legend.title=element_text(size=10), plot.title = element_text(size = 12,
face = "bold", hjust = 0.5),
axis.text.x = element_text(size = 9), axis.text.y = element_text(size = 9))
+
  guides(fill = guide_colorbar(barwidth = 6, barheight = 0.5,
title.position = "top", title.hjust = 0.5))

dev.off()

##----- Saving Annex tables -----

```

```
res_signif.int2 <- as.data.frame(res)
res_signif.int2$hgnc_symbol <- rownames(res_signif.int2)
apendiceA <- left_join(genes_int, res_signif.int2, by = "hgnc_symbol")
apendiceA$fold_change <- ifelse(apendiceA$log2FoldChange > 0,
2^apendiceA$log2FoldChange, -1 / (2 ^apendiceA$log2FoldChange))
write.xlsx(apendiceA, "results/ApendiceA.xlsx")
## Apendice B immquant
write.xlsx(vst_immquant_scores, "results/ApendiceB.xlsx")
```

```
> sessionInfo()
```

```
R version 4.0.0 (2020-04-24)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: windows 10 x64 (build 18363)
```

```
Matrix products: default
```

```
locale:
```

```
[1] LC_COLLATE=Portuguese_Brazil.1252 LC_CTYPE=Portuguese_Brazil.1252 LC_MONETA
RY=Portuguese_Brazil.1252
[4] LC_NUMERIC=C LC_TIME=Portuguese_Brazil.1252
```

```
attached base packages:
```

```
[1] grid stats4 parallel stats graphics grDevices utils datasets
methods base
```

```
other attached packages:
```

```
[1] EnhancedVolcano_1.6.0 ggrepel_0.8.2 ggcorrplot_0.1.3
ComplexHeatmap_2.4.2
[5] org.Hs.eg.db_3.11.4 enrichplot_1.8.1 clusterProfiler_3.16.0
EnsDb.Hsapiens.v86_2.99.0
[9] ensemblDb_2.12.1 AnnotationFilter_1.12.0 GO.db_3.11.4
vsn_3.56.0
[13] DESeq2_1.28.1 SummarizedExperiment_1.18.1 DelayedArray_0.14.0
matrixStats_0.56.0
[17] BiocParallel_1.22.0 GenomicFeatures_1.40.0 AnnotationDbi_1.50.0
GenomicRanges_1.40.0
[21] GenomeInfoDb_1.24.2 IRanges_2.22.2 S4Vectors_0.26.1
GEOquery_2.56.0
[25] Biobase_2.48.0 BiocGenerics_0.34.0 reshape2_1.4.4
xlsx_0.6.3
[29] dendextend_1.13.4 circlize_0.4.10 RColorBrewer_1.1-2
pheatmap_1.0.12
[33] forcats_0.5.0 stringr_1.4.0 dplyr_1.0.0
purrr_0.3.4
[37] readr_1.3.1 tidyr_1.1.0 tibble_3.0.1
ggplot2_3.3.2
[41] tidyverse_1.3.0 readxl_1.3.1 biomaRt_2.44.1
```

```
loaded via a namespace (and not attached):
```

```
[1] tidyselect_1.1.0 RSQLite_2.2.0 scatterpie_0.1.4 mu
nsell_0.5.0 preprocessCore_1.50.0
[6] withr_2.2.0 colorspace_1.4-1 GOsemSim_2.14.0 rs
studioapi_0.11 rJava_0.9-12
[11] DOSE_3.14.0 urltools_1.7.3 GenomeInfoDbData_1.2.3 po
lyclip_1.10-0 bit64_0.9-7
[16] farver_2.0.3 downloader_0.4 vctrs_0.3.1 ge
nerics_0.0.2 BiocFileCache_1.12.0
[21] R6_2.4.1 clue_0.3-57 graphlayouts_0.7.0 lo
cfit_1.5-9.4 bitops_1.0-6
[26] fgsea_1.14.0 gridGraphics_0.5-0 assertthat_0.2.1 sc
ales_1.1.1 ggraph_2.0.3
[31] gtable_0.3.0 affy_1.66.0 tidygraph_1.2.0 rl
ang_0.4.6 genefilter_1.70.0
[36] GlobalOptions_0.1.2 splines_4.0.0 rtracklayer_1.48.0 la
zyeval_0.2.2 broom_0.5.6
[41] europepmc_0.4 BiocManager_1.30.10 modelr_0.1.8 ba
ckports_1.1.8 qvalue_2.20.0 affyio_1.58.0 el
[46] tools_4.0.0 ggplotify_0.0.5
lipsis_0.3.1 ggribbles_0.5.2 plyr_1.8.6 progress_1.2.2 zl
[51] Rcpp_1.0.4.6 RCurl_1.98-1.2
ibbioc_1.34.0 openssl_1.4.1 GetoptLong_1.0.0 vi
[56] prettyunits_1.1.1 cowplot_1.0.0
ridis_0.5.1 cluster_2.1.0 fs_1.4.1 ma
[61] haven_2.3.1 data.table_1.12.8
gritr_1.5 triebeard_0.3.0 repress_0.3.0 Pr
[66] DO.db_2.9 hms_0.5.3
otGenerics_1.20.0 xtable_1.8-4 XML_3.99-0.3 gr
[71] xlsxjars_0.6.1 shape_1.4.4
idExtra_2.3 crayon_1.3.4 genepLOTter_1.66.0 lu
[76] compiler_4.0.0 DBI_1.1.0
bridate_1.7.9 dbplyr_1.4.4 MASS_7.3-51.6 ra
[81] tweenr_1.0.1 Matrix_1.2-18
ppdirs_0.3.1 igraph_1.2.5 pkgconfig_2.0.3 rv
[86] cli_2.0.2 GenomicAlignments_1.24.0
check_0.1.8
```

[91] xml2_1.3.2	annotate_1.66.0	xVector_0.28.0	rv
est_0.3.5	digest_0.6.25		
[96] Biostrings_2.56.0	cellranger_1.1.0	fastmatch_1.1-0	cu
r1_4.3	Rsamtools_2.4.0		
[101] rjson_0.2.20	lifecycle_0.2.0	nlme_3.1-148	js
onlite_1.6.1	viridisLite_0.3.0		
[106] askpass_1.1	limma_3.44.3	fansi_0.4.1	pi
llar_1.4.4	lattice_0.20-41		
[111] httr_1.4.1	survival_3.2-3	glue_1.4.1	pn
g_0.1-7	bit_1.1-15.2		
[116] ggforce_0.3.1	stringi_1.4.6	blob_1.2.1	me
moise_1.1.0			