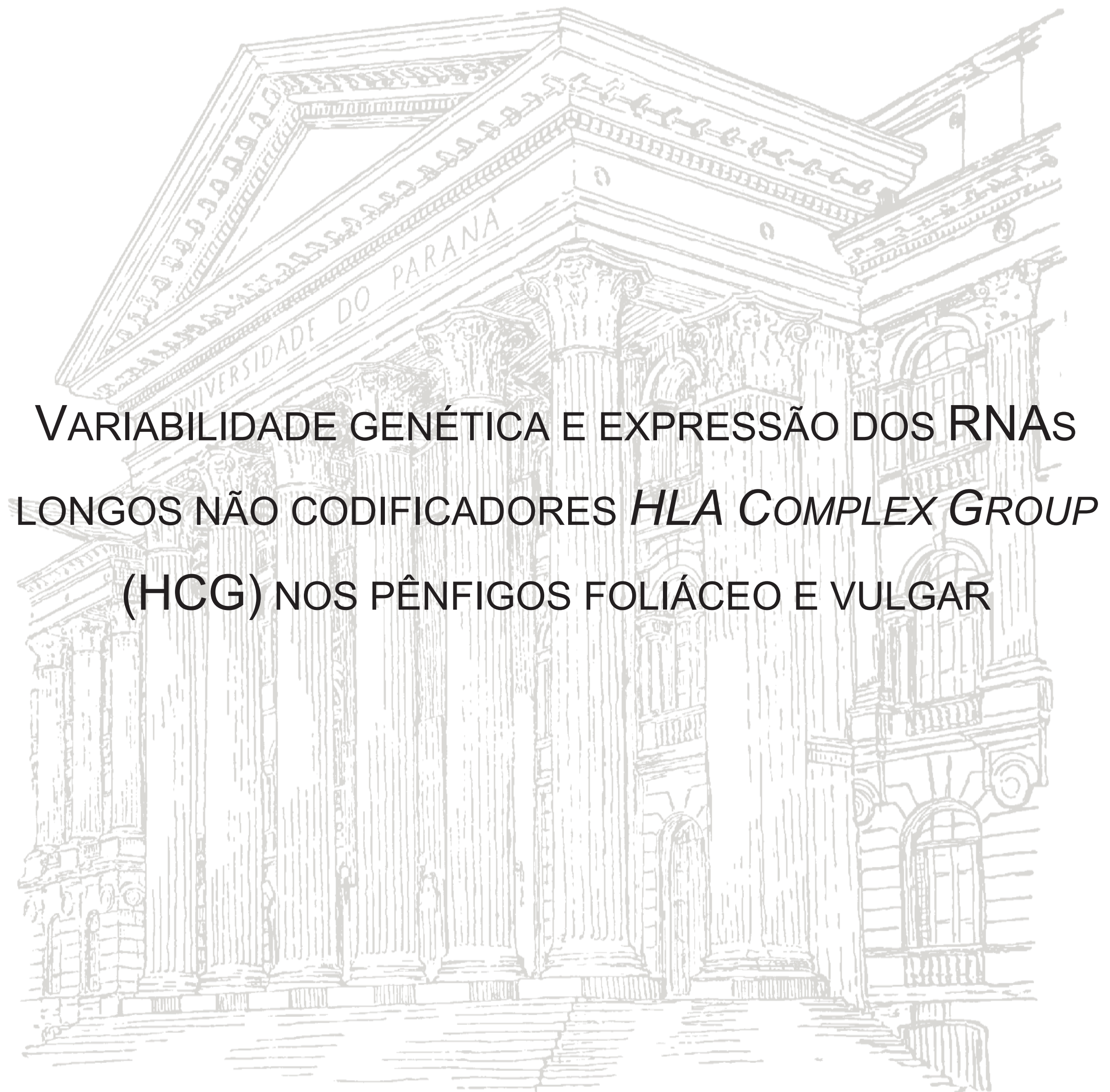
UNIVERSIDADE FEDERAL DO PARANÁ

AMANDA SALVIANO DA SILVA




# VARIABILIDADE GENÉTICA E EXPRESSÃO DOS RNAs LONGOS NÃO CODIFICADORES *HLA COMPLEX GROUP* (HCG) NOS PÊNFIGOS FOLIÁCEO E VULGAR

CURITIBA

2020

UNIVERSIDADE FEDERAL DO PARANÁ

AMANDA SALVIANO DA SILVA

# VARIABILIDADE GENÉTICA E EXPRESSÃO DOS RNAS LONGOS NÃO CODIFICADORES *HLA COMPLEX GROUP* (HCGS) NOS PÊNFIGOS FOLIÁCEO E VULGAR

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Genética, Setor de Ciências Biológicas, da Universidade Federal do Paraná.

Orientadora: Profª. Drª. Danielle Malheiros Ferreira
Coorientadores: Profª. Drª. Angelica Beate Winter Boldt e Dr. Rodrigo Coutinho de Almeida

CURITIBA

2020

Universidade Federal do Paraná. Sistema de Bibliotecas.
Biblioteca de Ciências Biológicas.
(Rosilei Vilas Boas – CRB/9-939).

Silva, Amanda Salviano da.
Variabilidade genética e expressão dos RNAs longos não codificadores HLA Complex Group (HCGs) nos pênfigos foliáceo e vulgar. / Amanda Salviano da Silva. – Curitiba, 2020.
168 f. : il.

Orientadora: Danielle Malheiros Ferreira.
Coorientadora: Angelica Beate Winter Boldt.
Coorientador: Rodrigo Coutinho de Almeida.

Tese (Doutorado) – Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Pênfigo. 2. Genética humana. 3. Autoimunidade. 4. Pele – Doenças – Aspectos genéticos. I. Título. II. Ferreira, Danielle Malheiros. III. Boldt, Angelica Beate Winter. IV. Almeida, Rodrigo Coutinho de. V. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

CDD (20. ed.) 616.5

MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIENCIAS BIOLOGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GENÉTICA - 40001016006P1

**ATA Nº134**

# ATA DE SESSÃO PÚBLICA DE DEFESA DE DOUTORADO PARA A OBTENÇÃO DO GRAU DE DOUTOR EM GENÉTICA

No dia vinte de março de dois mil e vinte às 14:00 horas, na sala Sala 65, Departamento de Genética, foram instaladas as atividades pertinentes ao rito de defesa de tese da doutoranda **AMANDA SALVIANO DA SILVA**, intitulada**: Variabilidade genética e expressão dos RNAs longos não codificadores HLA Complex Group (HCGs) nos pênfigos foliáceo e vulgar.**, sob orientação da Profa. Dra. DANIELLE MALHEIROS FERREIRA. A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação em GENÉTICA da Universidade Federal do Paraná, foi constituída pelos seguintes Membros: DANIELLE MALHEIROS FERREIRA (UNIVERSIDADE FEDERAL DO PARANÁ), VANESSA SANTOS SOTOMAIOR (PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ), WILSON ARAUJO DA SILVA JUNIOR (FACULDADE DE MEDICINA DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO), MARCIA REGINA PINCERATI (UNIVERSIDADE POSITIVO ), GABRIEL ADELMAN CIPOLLA (UNIVERSIDADE FEDERAL DO PARANÁ). A presidência iniciou os ritos definidos pelo Colegiado do Programa e, após exarados os pareceres dos membros do comitê examinador e da respectiva contra argumentação, ocorreu a leitura do parecer final da banca examinadora, que decidiu pela APROVAÇÃO. Este resultado deverá ser homologado pelo Colegiado do programa, mediante o atendimento de todas as indicações e correções solicitadas pela banca dentro dos prazos regimentais definidos pelo programa. A outorga de título de doutor está condicionada ao atendimento de todos os requisitos e prazos determinados no regimento do Programa de Pós-Graduação. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, DANIELLE MALHEIROS FERREIRA, lavrei a presente ata, que vai assinada por mim e pelos demais membros da Comissão Examinadora.

CURITIBA, 20 de Março de 2020.

Assinatura Eletrônica
03/04/2020 10:15:09.0
DANIELLE MALHEIROS FERREIRA
Presidente da Banca Examinadora

Assinatura Eletrônica
06/04/2020 14:44:56.0
VANESSA SANTOS SOTOMAIOR
Avaliador Externo (PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ)

Assinatura Eletrônica
20/04/2020 10:53:21.0
WILSON ARAUJO DA SILVA JUNIOR
Avaliador Externo (FACULDADE DE MEDICINA DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO)

Assinatura Eletrônica
12/05/2020 10:52:19.0
MARCIA REGINA PINCERATI
Avaliador Externo (UNIVERSIDADE POSITIVO )

Assinatura Eletrônica
16/04/2020 23:16:07.0
GABRIEL ADELMAN CIPOLLA
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIENCIAS BIOLOGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GENÉTICA -
40001016006P1

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **AMANDA SALVIANO DA SILVA** intitulada: **Variabilidade genética e expressão dos RNAs longos não codificadores HLA Complex Group (HCGs) nos pênfigos foliáceo e vulgar.**, sob orientação da Profa. Dra. DANIELLE MALHEIROS FERREIRA, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.
A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 20 de Março de 2020.

Assinatura Eletrônica
03/04/2020 10:15:09.0
DANIELLE MALHEIROS FERREIRA
Presidente da Banca Examinadora

Assinatura Eletrônica
06/04/2020 14:44:56.0
VANESSA SANTOS SOTOMAIOR
Avaliador Externo (PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ)

Assinatura Eletrônica
20/04/2020 10:53:21.0
WILSON ARAUJO DA SILVA JUNIOR
Avaliador Externo (FACULDADE DE MEDICINA DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO)

Assinatura Eletrônica
12/05/2020 10:52:19.0
MARCIA REGINA PINCERATI
Avaliador Externo (UNIVERSIDADE POSITIVO )

Assinatura Eletrônica
16/04/2020 23:16:07.0
GABRIEL ADELMAN CIPOLLA
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

## AGRADECIMENTOS

Como para tudo em minha vida, agradeço primeiramente aos meus queridos pais: *Maria Iva* e *Jivanildo*. A eles, devo tudo o que sou e tudo o que conquistei até então. Me sinto altamente privilegiada por fazer parte dessa família maravilhosa, que não mede esforços em me apoiar de todas as formas, e que vem me estimulando e dando forças durante toda a minha trajetória acadêmica. Sempre enxergarei os nomes dos meus pais estampados de alguma forma em cada um dos meus diplomas. E sempre estarei aliviada em causar orgulho nestas pessoas que também representam um grande orgulho para mim. Agradeço também aos meus irmãos (Yasmin, Rogério, Rodrigo e Márcio) e demais familiares, que também tem me apoiado e desejado o que há de melhor para mim. Agradeço à minha filhota Hathor, por sempre estar ali como uma "bolinha de pelo" esbanjando amor e carinho, sem me deixar sentir sozinha nem por um minuto sequer.

Agradeço à minha melhor amiga eterna, Aline Fontes (Nini). Desde sempre (e para sempre) me escutando, compreendendo e me fazendo enxergar a vida de forma mais leve. Com essa irmã de alma, sei que posso contar para o que der e vier. A Nini é praticamente PhD em levantar o meu astral e me fazer dar altas risadas, mesmo das coisas mais banais. Agradeço também a outros grandes amigos, com quem já contei diversas vezes e que já fizeram muitos dos meus dias mais divertidos: Camilla, Viridiana (Veri), Carlos (Death), Thiago (Satyr), entre outros. Agradeço ainda à Gi e ao Tiago, que apesar de situações naturais do cotidiano terem nos distanciado, estiveram muito ao meu lado durante todo o meu doutorado.

Agradeço à minha querida orientadora, Danielle Malheiros, por ter me acolhido como sua aluna de doutorado. Sempre agindo de forma profissional, ética e amigável, me apoiando na minha trajetória acadêmica e me fazendo crescer. Sempre disposta a ensinar, explicar e aprender em conjunto com todos os seus orientados. Dani, muito obrigada por toda a confiança e excelência. Foi um prazer ser sua aluna!

Agradeço à minha coorientadora Angelica Boldt, que desde o meu mestrado está ao meu lado me ensinando, incentivando e despertando um interesse quase que incontrolável pela ciência. Digamos que com a Angel, descobri um lado *workaholic* em mim (que apesar de estressante, tem me rendido bons frutos). Agradeço à professora Maria Luiza Petlz-Erler, que de certa forma, também tem me "coorientado" durante essa minha jornada. A professora Luiza sempre representou para mim aquela imagem de mulher sábia e competente, que eu tanto almejo ser (ou chegar perto de ser) no futuro. Agradeço ao meu coorientador Rodrigo Almeida, e também ao Danillo Augusto, Patrícia Savio e demais

professores e funcionários do Laboratório de Genética Molecular Humana (LGMH) e do departamento de Genética da UFPR. Agradeço ao professor Saleh Ibrahim e à Mareike Witte, que me supervisionaram durante meu período de doutorado sanduíche na Universidade de Lübeck, assim como aos professores Hauke Busch, Jennifer Hundt e outros pesquisadores desta instituição.

Agradeço aos meus colegas do LGMH, que mesmo também compartilhando de algum sofrimento científico (como coletas que não acabam mais, PCRs que não funcionam, *western blots* inespecíficos, rendimento baixo de RNAs, dados para analisar, disciplinas para concluir, artigos para ler, trabalhos para fazer, teses e dissertações para escrever, e assim por diante...), estão sempre dispostos a ajudar e conversar, mantendo um ambiente de trabalho saudável e agradável. Dentre estes, agradeço especialmente: à Fernanda Berti, que não é apenas a minha dupla perfeita para trabalhar, como também se tornou uma grande amiga; à Ticiana Farias, que com toda a sua energia e pró-atividade, sempre esteve disposta a ajudar no que fosse necessário, além de propiciar diversos momentos divertidos; à Valéria Bumiller, uma pessoa super doce e cuidadosa com quem me envolvi ativamente em diferentes trabalhos, e com quem morei por 6 meses em Lübeck (viramos quase irmãs!); ao Gabriel Cipolla, pela boa companhia, motivação e por sempre ajudar de forma tão atenciosa; à Sara Alves, minha companheira de lncRNAs; à Verónica Solís, outro membro do *team pênfigo*, que também sempre foi uma ótima companhia e parceira de trabalho.

Agradeço aos pacientes e demais indivíduos que doaram suas amostras para a realização deste estudo, bem como aos hospitais onde as coletas foram feitas. Por fim, agradeço aos órgãos de fomento, que financiaram este e tantos outros trabalhos científicos: CAPES, CNPq, Fundação Araucária e DAAD.

***"A tarefa não é tanto ver aquilo que ninguém viu,***
***mas pensar o que ninguém ainda pensou***
***sobre aquilo que todo mundo vê."***

*Arthur Schopenhauer.*

## RESUMO

O pênfigo é um grupo de doenças autoimunes de pele, caracterizado pelo auto-reconhecimento de antígenos desmossomais, como desmogleína-1 (DSG1, em pênfigo foliáceo – PF), e -3 (DSG3, em pênfigo vulgar – PV), levando à acantólise e bolhas epidermais. Diversos estudos já demostraram associações de variantes genéticas (alelos, haplótipos, ou polimorfismos de nucleotídeo único – SNPs) com a susceptibilidade ao PF e PV. Como para outras doenças autoimunes, as associações mais fortes são com alelos de antígenos leucocitários humanos (HLA) de classe II, além de outros genes do complexo principal de histocompatibilidade (MHC). Recentemente, variantes genéticas localizadas em RNAs longos não codificadores (lncRNAs) vem sendo associados com diferentes doenças autoimunes, incluindo o PF. Neste estudo pioneiro, foi investigada a associação entre SNPs de lncRNAs de uma família multigênica localizada na região do MHC, denominada *HLA Complex Group* (HCG), com a susceptibilidade ao PF e PV. Para tal, a distribuição de SNPs em 13 genes de lncRNAs HCG (genotipados em chips de microarranjo *Illumina*) foi analisada em PF endêmico (227 pacientes e 194 controles brasileiros) e PV (241 pacientes e 1188 controles alemães), por regressão logística multivariada. Nove SNPs foram significativamente associados ao PF (FDR $p<0,05$), dos quais 7 se localizam no gene de lncRNA *TSBP1-AS1* (que abriga o gene *HCG23*); um em *HCG27* e um em *HCG21*. Além disso, 55 SNPs foram associados com PV (FDR $p<0,01$), pertencendo aos genes *TSBP1-AS1* (38 SNPs), *HCG17* (6), *HCG27* (5), *HCG22* (3), *HCG20* (2) e *HCG26* (1). Para investigar se os genes mais associados também estão diferencialmente expressos no pênfigo, foram então quantificados os níveis transcricionais de *TSBP1*, *TSBP1-AS1*, *HCG23* e *HCG27* em células mononucleares do sangue periférico (PBMC) de pacientes e controles brasileiros. Foi verificado que a expressão de HCG27 está mais elevada em PBMC de pacientes com PF (p=0,035, log2 DE = 1,3), enquanto *TSBP1-AS1* esteve menos expresso em PBMC de pacientes com PV (p=0,029, DE = -1,29), em relação a indivíduos controle. Em concordância com estes resultados, as mesmas diferenças de expressão de *HCG27* (p=0,032, DE = 1,7) e *TSBP1-AS1* (p=0,032, DE = -0,99) também foram encontradas em ensaios funcionais com queratinócitos (linhagem celular HaCaT), após estes serem estimulados com anticorpos IgG de pacientes (5 PV e 5 PF) e controles (5) de uma coorte alemã. Além disso, níveis de mRNA de TSBP1 também

estiveram diminuídos em PBMCs de PF endêmico (p=0,042, DE = -2,14), apesar de não apresentarem diferença em queratinócitos representativos de PF esporádico. Em suma, foi demonstrado neste estudo que os lncRNAs HCGs estão associados com a susceptibilidade genética ao PF e PV, sendo *HCG27* e *TSBP1-AS1* diferencialmente expressos nas respectivas doenças, em diferentes amostras. Estes resultados indicam um papel dos lncRNAs HCGs na patogênese dos pênfigos, e encorajam a condução de mais estudos funcionais que visem elucidar o papel destes genes no mecanismo molecular que contribui com o desencadeamento da autoimunidade patológica.

**Palavras-chave:** *HLA Complex Group*, lncRNAs, pênfigo, susceptibilidade genética.

**ABSTRACT**

Pemphigus is a group of autoimmune skin diseases characterized by the autorecognition of desmosomal antigens, such as desmoglein-1 (DSG1, for pemphigus foliaceus - PF) and -3 (DSG3, for pemphigus vulgaris - PV), leading to acantholysis and epithelial blisters. Several studies have shown associations of genetic variants (alleles, haplotypes, or individual single nucleotide polymorphisms - SNPs) with PF and PV susceptibility, especially alleles of class II human leukocyte antigen (HLA) genes and other major histocompatibility complex (MHC) genes. Recently, genetic variants located in long non-coding RNAs (lncRNAs) have also been associated with many autoimmune diseases, including PF. In this pioneer study, it was investigated the association of genetic variants in lncRNAs of a multigenic family located in MHC region, classified as HLA complex group (HCG), with the susceptibility to PF and PV. To this end, the distribution of SNPs located in 13 HCG lncRNA genes (genotyped by Illumina microarrays) was analyzed in endemic PF (227 patients and 194 controls from Brazil) and in PV (241 patients and 1188 controls from Germany), applying multivariate logistic regression. We found 9 SNPs associated with endemic PF (FDR $p<0.05$): 7 located in the *TSBP1-AS1* lncRNA gene (which hosts *HCG23*); one in *HCG27* and one in *HCG21*. Moreover, 55 SNPs were found associated with PV (FDR $p<0.01$), overlapping the genes *TSBP1-AS1* (38 SNPs), *HCG17* (6), *HCG27* (5), *HCG22* (3), *HCG20* (2) and *HCG26* (1). To investigate if the most associated genes are differentially expressed in pemphigus, the transcriptional levels of *TSBP1*, *TSBP1-AS1*, *HCG23* and *HCG27* were quantified in peripheral blood mononuclear cells (PBMC) of Brazilian patients and controls. The *HCG27* was found upregulated in PBMC of endemic PF patients ($p=0.035$, log2 FC = 1.3), while *TSBP1-AS1* was downregulated in PBMC of PV patients ($p=0.029$, FC = -1.29), when compared to control subjects. Accordingly, the same differences were also found in functional assays with cultured immortalized keratinocytes (HaCaT cell line) stimulated with IgG antibodies of patients (5 PV and 5 PF) and controls (5) from Germany. Moreover, *TSBP1* mRNA levels were decreased in PBMCs of endemic PF patients ($p=0.042$, FC = -2.14), although no differences were found for this gene in keratinocytes representing sporadic PF. Taken together, it was demonstrated that HCG lncRNAs are associated with the genetic susceptibility for PF and PV, being *HCG27* and *TSBP1-AS1* also differentially expressed in the respective diseases, in

different samples. These results indicate a role of HCG lncRNAs in pemphigus pathogenesis, and encourage the conduction of further studies to elucidate the role of these genes in molecular mechanisms contributing to the development of pathological autoimmunity.

**Keywords:** HLA Complex Group, lncRNAs, pemphigus, genetic susceptibility.

## LISTA DE FIGURAS

FIGURA 1. Diferenças clínicas entre PV e PF. ........................................................................ 23

FIGURA 2. Manifestações clínicas de pacientes com PF e PV. .............................................. 24

FIGURA 3. Mapa de áreas brasileiras endêmicas para o PF. ................................................... 25

FIGURA 4. Classificação e mecanismos regulatórios dos lncRNAs na célula. ..........................36

FIGURA 5. Localização genômica dos genes HCG no braço curto do cromossomo 6. ........... 42

FIGURA 6. Região genômica do ENSG00000225914 (hg19 - chr6:32,222,417–32,361,468)... 44

## LISTA DE SIGLAS E ABREVIATURAS

3'UTR – região não traduzida 3'

AIBD - doenças bolhosas autoimunes

ANRIL – RNA não-codificador antisenso no *locus* INK4

AP4B1-AS1 – subunidade beta 1 do complexo AP-4 antisenso 1

C2 - componente 2 do complemento

C3 – componente 3 do complemento

C4 - componente 4 do complemento

C5AR1 – receptor 1 do componente do complemento C5A

C8A – componente 8 do complemento

C9 - componente 9 do complemento

cAMP - monofosfato cíclico de adenosina

CD1D – membro 1 grupo de diferenciação 1

CD4 – grupo de diferenciação 4

CD28- grupo de diferenciação 28

CD33 - grupo de diferenciação 33

CD40 – grupo de diferenciação 40

CD40L – ligante do grupo de diferenciação 40

CD59 – grupo de diferenciação 59

CD86 – grupo de diferenciação 86

CIITA – transativador do MHC classe II

CR1 – receptor do complemento 1

CR2 – receptor do complemento 2

CTLA4 – proteína citotóxica associada a célula T 4

DE – diferença de expressão

DNA – ácido desoxirribonucleico

DSG1 – desmogleína 1

DSG3 – desmogleína 3

EGFRK - receptor cinase do fator de crescimento epidérmico

eQTL – *locus* de traço quantitativo de expressão

FC – *fold change*

GAS5 - parada de crescimento específico 5

GWAS - Estudos de associação do genoma total

H19 – transcrito imprintado expresso

HCG – grupo do complexo HLA (HLA complex group)

HCP5 – pseudogene do complexo HLA 5

HIV – vírus da imunodeficiência humana
HLA – antígeno leucocitário humano
HOTAIR – transcrito RNA antisenso ao HOX
HSP27 - proteína de choque térmico 27
IFNG-AS1 – interferon-gama antisenso 1
IGFL – ligante do fator de crescimento semelhante à insulina
IgG – imunoglobulina G
IgM – imunoglobulina M
IL-6 – interleucina 6
IL-10 – interleucina 10
iRNA – RNA de interferência
ITGAM – integrina alfa M
ITGAX - integrina alfa X
JNK - cinase N-terminal c-Jun
KIR – receptor de células *killer* semelhante à imunoglobulina
KLRG1 - receptor G1 semelhante à lectina de células *killer*
LAIR - receptor associado a leucócitos semelhante à imunoglobulina
LCR - complexo de receptores leucocitários
LENG8 – membro 8 do grupo receptores leucocitários
LENG8-AS1 – membro 8 do grupo receptores leucocitários antisenso 1
LILRA – subfamília A de receptores leucocitários semelhantes à imunoglobulina
lnc-PREX1 - Proteína de permutador de Rac 1 dependente de fosfatidilinositol 3,4,5-trifosfato humano
lncRNA – RNA longo não codificador
lincRNA - RNA longo não codificador intergênico
MASP1 - serina protease associada à lectina de ligação à manose
MHC – complexo de histocompatibilidade principal
MIC – polipetídeo relacionado ao MHC classe I
mRNA – RNA mensageiro
miRNA – micro RNA
NAT – transcrito antisenso natural
ncRNA – RNA não codificador
NOD1 – proteína contendo domínio de oligomerização de ligação a nucleotídeos
ORF – região aberto de leitura
p38 MAPK – proteína cinase ativada por mitógeno p38
PDCD1 – proteína de morte celular programada 1
piRNA – piwi RNA

PF – pênfigo foliáceo
PFE – pênfigo foliáceo endêmico
PRINS – RNA associado à psoríase induzido por estresse
PSORS1C3 – candidato 3 à susceptibilidade à psoríase
PV – pênfigo vulgar
RNA – ácido ribonucleico
rRNA – RNA ribossomal
SLE - lúpus eritematoso sistêmico
snRNA – pequeno RNA nuclear
snoRNA – pequeno RNA nucleolar
SNP – polimorfismo de nucleotídeo único
ST18 – supressor tumorigênico 18
sQTL – *locus* de traço quantitativo de processamento
Th1 – célula T auxiliar 1
Th2 – célula T auxiliar 2
TNF - fator de necrose tumoral
TNFSF13B – membro 13b da superfamília do fator de necrose tumoral
tRNA – RNA transportador
TSBP1 – proteína básica expressa em testículo
TSBP1-AS1 - proteína básica expressa em testículo antisenso 1
TSS - sítio de início de transcrição
VSTM1 – proteína 1 contendo domínio transmembrana e V-set

## SUMÁRIO

**1. INTRODUÇÃO** ..... **20**

**2. REVISÃO DE LITERATURA** ..... **22**

2.1 O PÊNFIGO ..... 22

2.1.1 Aspectos clínicos ..... 22

2.1.2 Epidemiologia ..... 24

2.1.3 Diagnóstico e tratamento ..... 26

2.1.4 Etiopatogenia ..... 27

*2.1.4.1 Imunopatogênese* ..... 27

*2.1.4.2 Fatores genéticos e imunogenéticos* ..... 29

2.2 RNAS LONGOS NÃO CODIFICADORES ..... 33

2.2.1 LncRNAs: definição e função ..... 33

2.2.2 LncRNAs e doenças ..... 37

*2.2.2.1 lncRNAs em doenças autoimunes* ..... 37

*2.2.2.2 lncRNAs em pênfigo* ..... 39

2.3 GENES *HLA COMPLEX GROUP* (HCG) ..... 40

**3. HIPÓTESE E JUSTIFICATIVA** ..... **47**

**4. OBJETIVOS** ..... **486**

4.1 OBJETIVO GERAL ..... 48

4.2 OBJETIVOS ESPECÍFICOS: ..... 48

**5. CAPÍTULO I:** ..... **49**

**1. INTRODUCTION** ..... **52**

**2. MATERIAL AND METHODS** ..... **53**

***2.1 Human samples*** ..... 53

***2.2 DNA samples and Microarray genotyping*** ..... 54

***2.3 Selection of lncRNA candidates and data quality control*** ..... 55

*2.4 Computational analysis* ........ 55

*2.5 Cell culture* ........ 56

*2.6 Quantification of RNA levels* ........ 56

*2.7 Statistical analysis* ........ 57

**3. RESULTS** ........ **57**

*3.1 SNPs in HCG lncRNA genes are associated with PV* ........ 57

*3.2 SNPs in HCG lncRNA genes are associated with endemic PF* ........ 58

*3.3 Functional annotations suggest potential relevance of TSBP1-AS1 and HCG27 variability* .. 58

*3.4 TSBP1-AS1 expression is low in PV's PBMCs* ........ 59

*3.5 HCG27 is overexpressed in EPF PBMCs* ........ 60

*3.6 TSBP1-AS1 and HCG27 are also differentially expressed in PV and PF IgG-stimulated keratinocytes* ........ 60

**4. DISCUSSION** ........ **61**

**REFERENCES** ........ **65**

***FIGURES AND TABLES*** ........ **71**

**6. CAPÍTULO II (Manuscrito de qualificação)** ........ **101**

**1. INTRODUCTION** ........ **105**

**2. MATERIAL AND METHODS** ........ **107**

2.1 Study populations and samples ........ 107

2.2 Microarray genotyping ........ 108

2.3 Selection of lncRNA candidates ........ 108

2.4 Genotyping data analysis ........ 108

2.5 In silico analysis ........ 109

2.6 Evaluation of lncRNA genotyping in sporadic PF ........ 109

2.7 Quantification of lncRNA relative expression levels ........ 110

**3. RESULTS** ........ **111**

3.1 SNPs in immune-related lncRNA genes are associated with endemic PF ............ 111
3.2 Computational and in silico results .................................................................. 111
3.3 LINC-PINT*rs10228040 is associated with sporadic PF ....................................... 112
3.4 LY86-AS1*rs12192707 marks differential expression........................................... 113
**4. DISCUSSION ........................................................................................................ 114**
**REFERENCES ........................................................................................................ 118**
***Figure Captions*** **........................................................................................................ 123**
**7. CONCLUSÃO FINAL ........................................................................................ 156**
**8. REFERÊNCIAS ..................................................................................................... 156**
**9. TRABALHO ANEXO (Artigo de revisão de literatura publicado)......................................... 169**

# 1. INTRODUÇÃO

O pênfigo é um grupo de doenças bolhosas autoimunes, caracterizado pela ação de autoanticorpos que reconhecem antígenos de adesão na superfície dos desmossomos, levando à ruptura de ligações intercelulares. No pênfigo foliáceo (PF), o principal antígeno reconhecido pelos anticorpos patogênicos é a desmogleína 1 (DSG1), enquanto que em pênfigo vulgar, o principal autoantígeno é a desmogleína 3 (DSG3). Este processo ocasiona a perda de adesão entre os queratinócitos (acantólise) e se manifesta través da formação de bolhas nas camadas superficiais da epiderme, com acometimento localizado ou generalizado. Apesar de ser uma doença rara, o PF é endêmico em certas regiões rurais brasileiras, onde é popularmente conhecido como “fogo selvagem”, devido à forte sensação de ardor na pele. Ainda não estão bem elucidados os mecanismos patológicos que desencadeiam o pênfigo, mas sabe-se que se trata de uma patologia de caráter multifatorial, onde fatores genéticos, imunológicos e ambientais estão envolvidos.

Há diversos genes associados com a susceptibilidade genética e de expressão em pênfigo, com grande destaque para genes de HLA classe II, entre outros na região do MHC. Entretanto, esta região abriga também uma família multigênica denominada *HLA complex group* (HCG), que inclui genes que são expressos em RNAs longos não codificadores (lncRNAs). Os lncRNAs são transcritos com mais de 200 nucleotídeos e sem capacidade codificadora de proteína, envolvidos em diferentes processos regulatórios transcricionais e pós-transcricionais. Devido a seus diversos mecanismos funcionais, os lncRNAs têm sido cada vez mais investigados em processos fisiológicos e patológicos.

São escassos ainda os estudos investigando a influência de lncRNAs na susceptibilidade ao PF e, até o momento, ausentes em PV. Considerando-se mais especificamente os HCGs, embora alguns destes genes já tenham sido previamente associados a doenças imuno-relacionadas, não há ainda qualquer estudo com os pênfigos. Portanto, tendo em vista a importância da região genômica e função dos lncRNAs a hipótese deste trabalho é a de que variantes de genes de lncRNAs HCGs possam estar associadas com a susceptibilidade aos pênfigos. Sendo assim, o objetivo deste estudo foi investigar, através de estudo caso-controle e de avaliação da

expressão gênica, se a variabilidade genética e os níveis de expressão de lncRNAs da família dos HCGs aqui selecionados, estão associadas com os pênfigos foliáceo e vulgar.

Este trabalho está organizado da seguinte maneira: as seções de revisão de literatura, hipótese e justificativa, objetivos e conclusão final seguem o modelo de tese de acordo com normas da UFPR; já as seções de material e métodos, resultados e discussão seguem o modelo de artigo científico permitido pelo Programa de Pós-Graduação em Genética.

# 2. REVISÃO DE LITERATURA

## 2.1 O PÊNFIGO

### 2.1.1 Aspectos clínicos

O pênfigo é um grupo de doenças que se caracteriza pela presença de bolhas e erosões na região cutânea e/ou mucosa, causados por autoanticorpos antiepidermais. Estas condições, apesar de em sua maioria raras, são bastante graves e fazem parte dos quatro principais grupos de doenças bolhosas autoimunes (AIBD) (DIAZ et al., 1989 ; SINHA, 2011; SCHMIDT, KASPERKIEWICZ & JOLY, 2019).

Nos pênfigos, os autoanticorpos patogênicos reconhecem antígenos de superfície dos queratinócitos, presentes na região do desmossomo e com função de comunicação intercelular. Dentre estas importantes glicoproteínas de adesão, o papel das desmogleínas é bem estabelecido nos mecanismos patológicos do pênfigo. As desmogleínas medeiam a interação com proteínas intracelulares, ligando o citoesqueleto de queratina com a superfície celular, permitindo força e resistência ao estresse mecânico sofrido pelo epitélio. A ruptura dessas ligações intercelulares ocasiona a perda de adesão entre os queratinócitos e a formação de bolhas intraepidermais, processo este conhecido como acantólise, característico no pênfigo (TAKEICHI, 1991; HERTL, EMING & VELDMAN, 2006; KASPERKIEWICZ et al., 2017; POLLMANN et al., 2018; SPINDLER et al., 2018; SCHMIDT, KASPERKIEWICZ & JOLY, 2019).

Existem dois tipos clássicos de pênfigo: o pênfigo vulgar (PV) e o pênfigo foliáceo (PF), que diferem entre si de acordo com a região da pele afetada, sintomas e antígenos alvo. O principal autoantígeno associado ao PV é a desmogleína 3 (DSG3) (FIGURA 1A), uma proteína presente em camadas mais profundas da epiderme, ocasionando assim o desenvolvimento de bolhas abaixo da camada basal, tanto na pele quanto em mucosas. Já o PF está associado com a presença de autoanticorpos contra a desmogleína 1 (DSG1) (FIGURA 1B), cuja maior expressão ocorre na lâmina superficial da epiderme, causando bolhas superficiais logo abaixo do estrato

córneo, e sendo raras ou ausentes em regiões mucosas (BYSTRYN & RUDOLPH, 2005; AMAGAI et al., 1999; MAHONEY et al., 1999; NISIHARA et al., 2003; KITAJIMA & AOYAMA, 2007; SCHMIDT, KASPERKIEWICZ & JOLY, 2019).

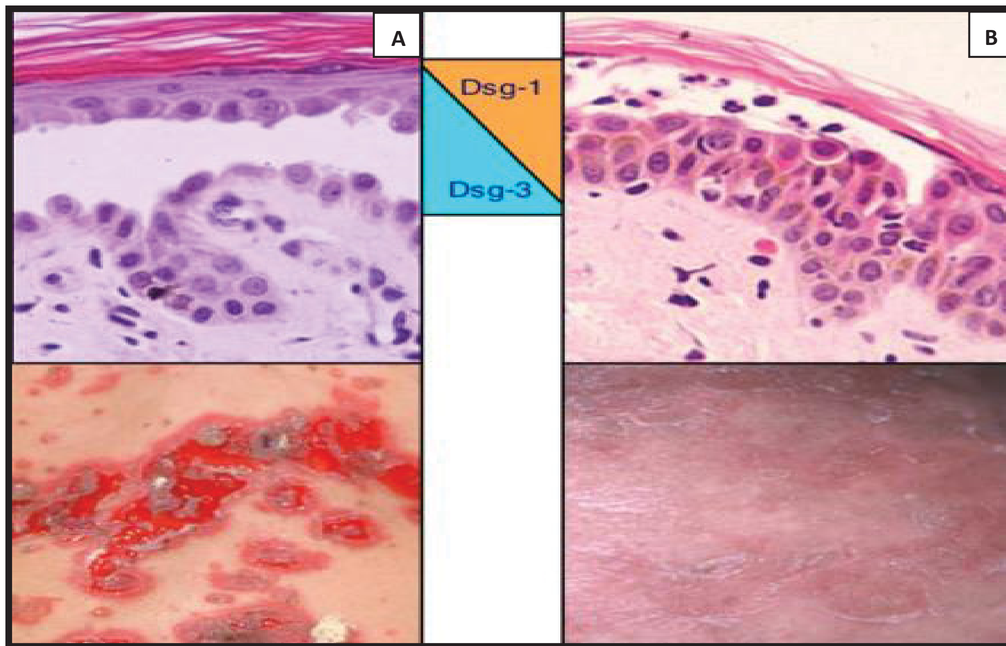



FIGURA 1. Diferenças histológicas e clínicas entre PV e PF. **(A)** O PV caracteriza-se pelo auto-reconhecimento de DSG3, abundante nas camadas mais profundas da epiderme, causando acantólise e formação de bolhas profundas. **(B)** O PF caracteriza-se pelo auto-reconhecimento de DSG1, mais abundante em camadas mais superficiais, causando acantólise e formação de bolhas na superfície da epiderme. Fonte: KITAJIMA e AOYAMA (2007).

Em pacientes de PF com a doença ativa (FIGURA 2 A-C), as lesões primárias se caracterizam por pequenas bolhas hiperpigmentadas situadas em áreas seborreicas, predominantemente na região da cabeça (em especial na face e couro cabeludo), pescoço e tronco superior. São raras as lesões na cavidade oral, porém estas podem ocorrer superficialmente. Em pacientes com a doença mais agressiva, as lesões se apresentam de forma generalizada, estendendo-se para o restante do tronco, braços e pernas. As bolhas se rompem facilmente, formando crostas, e são agravadas por diversos fatores, como a exposição excessiva ao sol. Em geral, estes pacientes sentem prurido, dor e sensação de queimação, pela excessiva descamação da pele (DIAZ et al., 1989; BYSTRYN & RUDOLPH, 2005; CULTON et al., 2008). Já em pacientes de PV (FIGURA 2 D-F), as lesões são mais profundas e geralmente se

originam na cavidade oral, podendo ainda acometer outras regiões mucosas, como esôfago, faringe e mucosas genitais (SCHMIDT, KASPERKIEWICZ & JOLY, 2019).

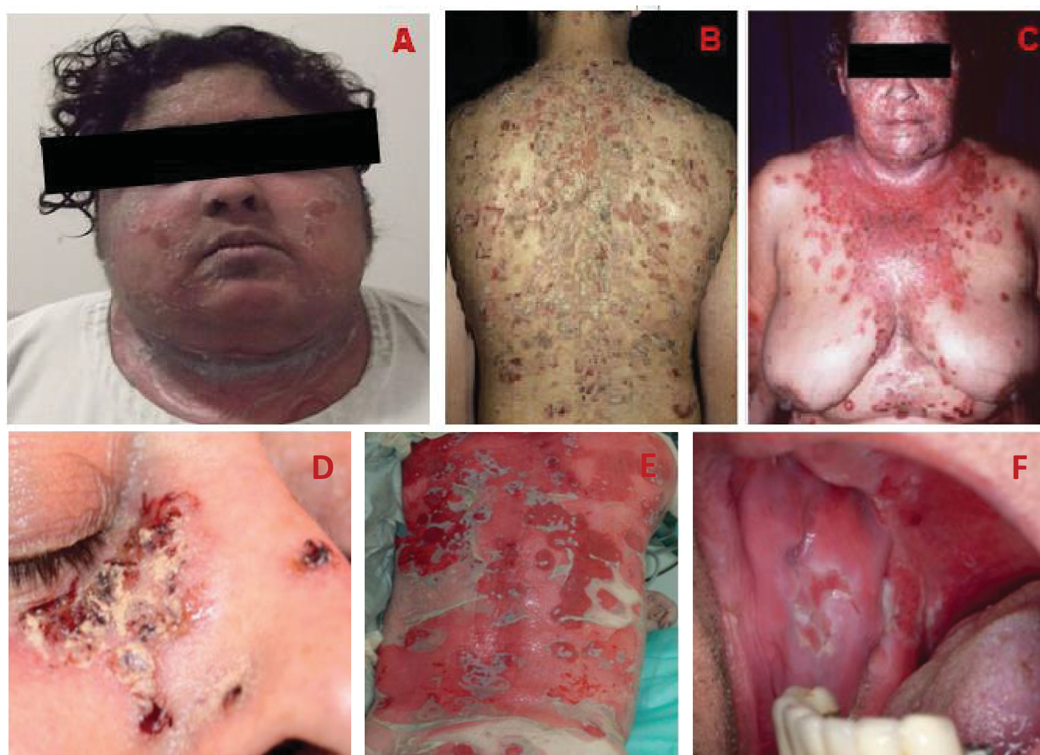



FIGURA 2. Manifestações clínicas de pacientes com PF e PV. **(A)** Bolhas localizadas na região da face e pescoço de adolescente com PF endêmico. O inchaço pelo corpo é um efeito característico da corticoterapia. Fonte: acervo LGMH/UFPR. **(B e C)** Pacientes de PF com bolhas generalizadas por todo o corpo. Fonte: adaptado de DIAZ et al. (1989) e de CULTON et al. (2008). **(D, E e F)** Pacientes de PV, com lesões na face, costas e mucosa oral. Fonte: SCHMIDT, KASPERKIEWICZ & JOLY (2019).

### 2.1.2 Epidemiologia

Os Pênfigos possuem incidência mundial de 0,75-5 casos/milhão por ano, variando consideravelmente de acordo com as formas da doença entre algumas populações (BYSTRYN & RUDOLPH, 2005), sendo o PV a forma mais frequente, acometendo cerca de 70% dos casos (JOLY & LITROWSKI, 2011). A incidência de PV varia de <1 casos/milhão por ano em alguns países da Europa, até 16 casos/milhão por ano nos Estados Unidos e 32 casos/milhão em Israel, afetando principalmente indivíduos de origem judaica (judeus *Ashkenazi*, por exemplo) (KRIDIN, 2018). Já o PF ocorre de forma esporádica na maior parte das populações, com incidência de 0.5–1 casos/milhão por ano na Europa ocidental (PAYNE & STANLEY,

2012). Há ainda outra forma de PF, o pênfigo foliáceo endêmico (PFE), popularmente conhecido como “fogo selvagem”, que difere da forma esporádica por possuir caráter endêmico em algumas regiões rurais do Brasil, Colômbia, Peru e Tunísia. No Brasil, o PFE acomete áreas de clima subtropical, principalmente nas regiões centro-oeste e sudeste (FIGURA 3), atingindo sua maior incidência mundial de aproximadamente 3,4% em uma comunidade indígena Terena (reserva Limão-Verde) no interior do Mato Grosso do Sul (BYSTRYN & RUDOLPH, 2005; DIAZ et al., 1989; MEYERA & MISERY, 2010; DIAZ et al., 1989 [2]; HANS-FILHO et al., 1999).

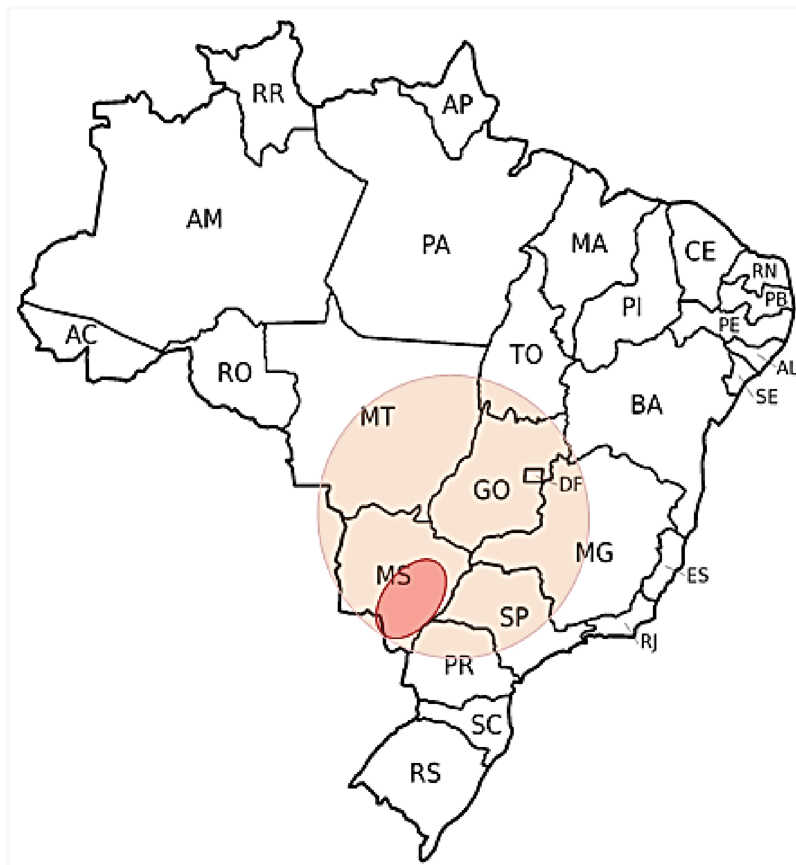



FIGURA 3. Mapa de áreas brasileiras endêmicas para o PF. Em laranja, região endêmica para PF no Brasil, com destaque para região com maior número de casos reportados para a doença (em vermelho). Fonte: a autora (2020).

A manifestação dos pênfigos parece apresentar predominância para o sexo feminino em algumas populações (SALVIANO-SILVA et al., 2017; KRIDIN, 2018; SCHMIDT, KASPERKIEWICZ & JOLY, 2019). Em contraste com as formas raras de pênfigo, que geralmente se manifestam entre 45 e 65 anos de idade (SCHMIDT,

KASPERKIEWICZ & JOLY, 2019), o PFE apresenta pico de incidência entre a segunda e terceira décadas de vida (AOKI et al., 2011). Além dos diferentes fatores genéticos associados com a doença nestas populações (ver tópico 2.1.4.2 desta tese), a etiologia do PFE depende da exposição a potentes antígenos ambientais ainda pouco conhecidos, acometendo principalmente habitantes de áreas rurais, próximas a rios e com alta frequência de insetos hematófagos, dos quais, algumas espécies são suspeitas de participar do desencadeamento da doença. Dentre estes, destacam-se os da família dos simulídeos, popularmente conhecidos como "borrachudos", especialmente a espécie *Simulium nigrimanum* (CULTON et al., 2008), cujo sialotranscriptoma já foi verificado com mais de 30 famílias de proteínas passíveis de reação com o soro de pacientes de PFE (AOKI et al., 2011). Além dos simulídeos, também há uma forte suspeita do envolvimento da espécie *Lutzomyia longipalpis* no PFE. Quian e colaboradores, em estudos com este flebotomíneo, descobriram que as proteínas com a sequência de aminoácidos SGLL, presentes em suas glândulas salivares, eram antigênicas para anticorpos anti-desmogleína IgG4 e IgE do soro de pacientes com PF (QIAN et al., 2012).

### 2.1.3 Diagnóstico e tratamento

O diagnóstico de pênfigo é realizado a partir da avaliação das manifestações clínicas (erosões e bolhas na pele); técnicas histológicas (presença de acantólise epidermal); e imunológicas, por ensaios de imunofluorescência indireta no soro, ou por imunofluorescência direta em biópsias de pele, avaliando depósitos de anticorpos IgG contra DSG1 e/ou 3, e/ou ainda moléculas de C3 (*complement component 3*) nas lesões (CAMPBELL et al., 2001; BYSTRYN e RUDOLPH, 2005; SCHMIDT, KASPERKIEWICZ & JOLY, 2019).

O tratamento atual se faz pelo uso de corticosteroides tópicos, orais ou ainda intralesionais. Podem ser administrados coadjuvantes em associação, como imunossupressores, anti-inflamatórios e antibióticos. Porém, essas opções terapêuticas não são específicas para pênfigo e variam de paciente para paciente, além da ocorrência de diversos efeitos colaterais (CAMPBELL et al., 2001; SCHMIDT, KASPERKIEWICZ & JOLY, 2019) e infecções extra-cutâneas por patógenos oportunistas (BYSTRYN e RUDOLPH, 2005; KORMAN, 1988).

#### 2.1.4 Etiopatogenia

A etiopatogenia do pênfigo é multifatorial. O aparecimento depende da qualidade da resposta do hospedeiro frente a fatores ambientais e/ou agentes desencadeantes, qualidade esta que depende da sua susceptibilidade genética à doença. Após o estímulo ambiental em portadores de alelos de HLA (*human leukocyte antigens*) classe II de alto risco, certos peptídeos de desmogleína podem ser apresentados a células T $CD4^+$ autorreativas. A ativação destas células T leva a produção de autoanticorpos específicos para DSG1 e/ou DSG3 a partir de células B, ocasionando a destruição dos desmossomos, perda de adesão entre queratinócitos e acantólise, com consequente formação de bolhas (CULTON et al., 2008; SCHMIDT, KASPERKIEWICZ & JOLY, 2019).

##### *2.1.4.1 Imunopatogênese*

Diferentes processos participam da imunopatogênese do pênfigo. Dentre estes, é possível que os autoanticorpos do pênfigo induzam a ativação de proteinases, principalmente o plasminogênio ativado. Este se transforma em plasmina, processo este que levaria à lise de substâncias intercelulares, rompendo as pontes entre as desmogleínas. Dessa forma, os demossomos assumiriam um aspecto arredondado, perdendo a sua adesão (AMORMINO e BARBOSA, 2010; OLIVEIRA-JR, 2012). Contudo, a participação do sistema plasmina/plasminogênio é bastante controversa no processo acantolítico, de forma que o mesmo foi demonstrado por não ser necessário para a ativação da acantólise em camundongos (MAHONEY, WANG & STANLEY, 1999).

Outro processo possivelmente associado com a acantólise é a ativação exacerbada do sistema complemento, que ocorre devido à ligação patogênica antígeno-anticorpo, levando à deposição de seus componentes líticos nos espaços intercelulares da epiderme (JORDON et al., 1974; KAWANA et al., 1988, 1989; MESSIAS-REASON et al. 1989, 2008, 2011), sendo a deposição do componente C3 já utilizada a nível diagnóstico (SCHMIDT, KASPERKIEWICZ & JOLY, 2019). Além

disso, genes da cascata do complemento já foram associados com o PFE (SALVIANO-SILVA et al., 2017; BUMILLER-BINI et al., 2018; OLIVEIRA, L.C. et al., 2019).

Dentre outras hipóteses que expliquem a perda de adesão entre os queratinócitos, há ainda a evidência de que anticorpos IgG anti-desmogleína possam ativar diferentes eventos de sinalização celular, por se ligarem a distintos receptores de superfície. Destes, já foram reportadas as participações das vias de receptor cinase do fator de crescimento epidérmico (EGFRK), monofosfato cíclico de adenosina (cAMP), proteína de choque térmico 27 (HSP27), proteína cinase ativada por mitógeno p38 (p38 MAPK), cinase N-terminal c-Jun (JNK) entre outros receptores de tirosina cinase no processo acantolítico. Isso levaria a ativação de vias de morte celular em queratinócitos, como a apoptose e oncose, em um processo denominado apoptólise (Grando, 2011). Por fim, além da apoptose, sugere-se ainda que a acantólise também tenha relação com outras vias de morte celular, como a via imunogênica, para qual já foram encontradas diversas evidências em termos de associação genética (BUMILLER-BINI et al., 2019). Além desta, também se sugere o envolvimento de vias da piroptose, necroptose e outras (BUMILLER-BINI et al., 2019; FRUSIC-ZLOTKIN et al., 2005; PUVIANI et al., 2003; DEYHIMI & ALISHAHI et al., 2018; LEE et al., 2009).

As lesões epidermais possuem um elevado infiltrado inflamatório, com predominância de células T $CD4^{+}$, previamente formadas em órgãos linfoides primários e necessárias para a ativação dos linfócitos B (produtores de anticorpos anti-Dsg1/3) nos linfonodos. Possuem resposta do tipo Th2, que leva a uma maior produção de anticorpos IgG4, em relação a IgG1 (CAPRONI et al., 2001; HERTL, EMING & VELDMAN, 2006; MALHEIROS, 2009; PEREIRA et al., 2004; CULTON et al., 2008). Os autoanticorpos anti-desmogleína da subclasse IgG1 são encontrada em maiores níveis no soro de pacientes de pênfigo em estágio pré-clínico ou em remissão, ou ainda de indivíduos saudáveis que habitam regiões de endemia para o PFE. Este fato demonstra que o IgG1 não é, por si só, suficiente para desencadear a doença, sendo portanto, não patogênico. Já a subclasse IgG4 de anticorpos anti-desmogleína é encontrada quase que exclusivamente em indivíduos com a doença ativa, demonstrando haver uma importância dessa imunoglobulina na patogênese do

pênfigo. Dessa forma, o agente desencadeador parece estimular a produção de anticorpos IgG1, porém a doença só se manifesta em indivíduos que possuam uma susceptibilidade genética que permita a produção de anticorpos específicos do tipo IgG4, levando à resposta autoimune patogênica (BYSTRYN e RUDOLPH, 2005; CULTON et al., 2008; AOKI et al., 2004).

Os anticorpos IgM dirigidos contra a desmogleína parecem não possuir relevância funcional no pênfigo, contudo, estão presentes no soro de pacientes de PFE e de muitos indivíduos saudáveis residentes em áreas endêmicas, apoiando a hipótese da sensibilização imunológica de por um vetor em áreas endêmicas (CULTON et al., 2008, QIAN et al., 2012). Os níveis de IgM tendem a diminuir quando o indivíduo migra para regiões urbanas não endêmicas (AOKI et al., 2011), e podem distinguir clinicamente pacientes de PFE e de PF esporádico (DIAZ et al., 2008, QIAN et al., 2012). Além de autoanticorpos anti-desmogleína IgG4 e IgM, pacientes de PFE também possuem níveis elevados de IgE. Apesar do mecanismo envolvido no aumento dos níveis destes autoanticorpos serem ainda desconhecidos (QIAN et al., 2012), já foi demonstrado que maiores níveis de IgE correlacionam-se com a atividade da doença em pacientes de PV (NAGEL et al., 2010).

#### *2.1.4.2 Fatores genéticos e imunogenéticos*

Diversos estudos já demonstraram que a variabilidade genética (principalmente de nucleotídeos de polimorfismo único - SNPs) e diferenças de expressão gênica estão implicadas com a susceptibilidade diferencial ao PF e PV. Dentre os genes mais associados com o pênfigo, destacam-se certos alelos da família dos antígenos leucocitários humanos (HLA, do inglês *Human Leukocyte Antigens*) de classe II, especialmente *HLA-DRB1* e *HLA-DQB1*. Os genes HLA estão localizados na região genômica do complexo principal de histocompatibilidade (MHC, do inglês *Major Histocompatibility Complex*), e estão envolvidos na modulação da susceptibilidade e resistência a doenças autoimunes (TRON et al., 2005). Nos pênfigos, propõe-se que quando um indivíduo geneticamente susceptível entra em contato com o agente desencadeador da resposta imune, há um aumento da produção de interleucinas, que estimulam uma resposta Th2 e induzem a expressão de HLA classe II nas

membranas dos queratinócitos, ativando a produção de autoanticorpos pela exposição de peptídeos de desmogleína (TRON et al., 2005; SAGI et al., 2008).

Em PV, a frequência dos alelos de *HLA DRB1*04, DRB1*08, DRB1*14, DQB1*03:02* e *DQB1*05:03* é significativamente maior em pacientes, em contraste com a menor frequência de *DRB1*07, DRB1*15, DQB1*03:03, DQB1*02, DRB1*03, DQB1*05:01* e *DQB1*06:01* (YAN, WANG & ZENG, 2012; LI et al., 2018, PETZL-ERLER, 2020). Dentre estes, *DRB1*04:02* e *DQB1*05:03* são os principais alelos de risco em diferentes populações, sendo expressos na maioria dos pacientes de PV (VODO, SARIG, & SPRECHER, 2018, PETZL-ERLER, 2020).

De forma similar, pacientes de PF também apresentam maior frequência de *DRB1*04* e *DRB1*14*, e uma menor frequência de *DRB1*07* (PETZL-ERLER & SANTAMARIA, 1989; LOMBARDI et al., 1999; PAVONI et al., 2003, PETZL-ERLER, 2020). Destes, *DRB1*04* é o alelo mais associado com a susceptibilidade para PF esporádico e PFE (PAVONI et al., 2003; MORAES et al., 1997; DE SENA et al., 2018, PETZL-ERLER, 2020), sendo que em PFE, portadores de *DRB1*04* apresentam 14 vezes mais chance de desenvolver a doença, em comparação a indivíduos não portadores deste alelo (MORAES et al., 1997). Já o alelo de risco para PV *DRB1*04:02* não está associado com as formas PF (LOMBARDI et al., 1999; BROCHADO et al., 2016, PETZL-ERLER, 2020).

Além disso, enquanto os alelos *DQB1*05* e *DRB1*16* também se apresentam como alelos de risco para PFE, os mesmos não se encontram associados com PF esporádico na população holandesa (DE SENA et al., 2018). Considerando que *DQB1*05:03* também é alelo de risco para PF na Itália (LOMBARDI et al., 1999), e que há um forte desequilíbrio de ligação entre os genes *HLA-DQB1*05* e *-DRB1*16*, a presença de associação destes alelos pode ser explicável em regiões brasileiras previamente colonizadas por italianos, sobretudo em São Paulo (DE SENA et al., 2018). Diferenças de susceptibilidade genética para PFE são encontradas ainda em ameríndios brasileiros. É o caso da população Terena, a qual o alelo HLA de alto risco é o *DRB1*04:04*, além de uma associação adicional de *DQB1*03:02* com a susceptibilidade à doença. Enquanto isso, o alelo *DRB1*08:02* (de alto risco para PV e baixo risco para ambas as formas de PF, em indivíduos de ascendência

predominantemente europeia) é considerado protetor para o PFE nesta população ameríndia (PETZL-ERLER, 2020).

Ainda dentro ou nas proximidades da região genômica do MHC, também já foram associados com o pênfigo, alguns genes de HLA classe I (VODO, SARIG, & SPRECHER, 2018; PETZL-ERLER e SANTAMARIA, 1989; AUGUSTO et al., 2012; BROCHADO et al., 2016) e genes não pertencentes à família dos HLA, como *TNF* (*Tumor Necrosis Factor*), *LTA* (*Lymphotoxin alpha*) e *TAP2* (*Transporter Antigen Peptide 2*) (TRON et al., 2005; VODO, SARIG, & SPRECHER, 2018).

Diversos genes não pertencentes ao MHC também vem sendo associados com o pênfigo em algumas populações, ao nível de variabilidade genética e/ou expressão gênica. Dentre estes, destacam-se: genes *KIR* (*Killer-cell Immunoglobulin-like Receptor*) (AUGUSTO et al., 2012); interleucinas, como *IL-6* e *IL-10* (PEREIRA et al., 2004; MOSAAD et al., 2012; FELICIANI et al., 2000; EBERHARD et al., 2005); reguladores de linfócitos T, como *CD28, CD86* e *CTLA4* (*cytotoxic T-lymphocyte-associated protein 4*) (DALLA-COSTA ET AL., 2010; PAVONI et al., 2006; TANASILOVIC et al., 2017); reguladores de linfócitos B, como *CD40, CD40L* e *BLYS* (*B Lymphocyte Stimulator*) (MALHEIROS & PETZL-ERLER, 2009); reguladores da resposta imune, como *CIITA* (*Class II MHC Transactivator*), *PDCD1* (*Programmed Cell Death Protein 1*) e *ST18* (*T18 C2H2C-type zinc finger transcription factor*) (PIOVEZAN & PETZL-ERLER, 2013; BRAUN-PRADO & PETZL-ERLER, 2007; SARIG et al., 2012); genes do complexo de receptores leucocitários, como os *LAIR1/2* (*Leukocyte-Associated Inhibitory Receptors 1* and *2*), *LILRB1/2 (Leukocyte Immunoglobulin Like Receptor B 1* and *2*), *LILRA3/4 (Leukocyte Immunoglobulin Like Receptor A 3* and *4*), *LENG8* (*leukocyte receptor cluster member 8*) e *VSTM1* (*V-Set and Transmembrane domain-containing protein 1*) (CAMARGO et al., 2016; FARIAS et al., 2018); e também diversos genes do sistema complemento (SALVIANO-SILVA et al., 2017; BUMILLER-BINI et al., 2018; OLIVEIRA et al., 2019). Além destes, também estão associados ao PF genes envolvidos em diferentes vias de morte celular (BUMILLER-BINI et al., 2019) e em mecanismos epigenéticos (SPADONI et al., 2019), entre outros.

Em nível transcriptômico, diversos genes foram encontrados diferencialmente expressos em biópsias de lesões de PFE, assim como em células T $CD4^+$ de

pacientes com lesões localizadas e generalizadas. Dentre estes, destacam-se genes envolvidos na apresentação de antígenos lipídicos, como o CD1D; na adesão e migração de linfócitos, como o CD33, na proliferação e estimulação linfocitária, como o TNFSF13B (BLYS); e na sinalização de MAPKs, alterando especialmente a via da p38 MAPK (MALHEIROS et al., 2014).

Além dos codificadores de proteínas, alguns genes de RNAs não codificadores (ncRNAs), como os microRNAs (miRNAs) e os RNAs longos não codificadores (lncRNAs), também foram recentemente associados com PFE. Os miRNAs são reguladores pós-transcricionais de aproximadamente 22 nucleotídeos, que se ligam à região 3'UTR do RNA mensageiro (mRNA), levando à sua degradação ou inibindo sua tradução. Em um estudo com miRNAs, observou-se que polimorfismos na região 3'UTR do gene *KLRG1* (*Killer cell Lectin like Receptor G1*) influenciam na atividade de ligação do miRNA miR-584-5p, portanto alterando a expressão de KLRG1, o que poderia contribuir para a autoimunidade do PF (CIPOLLA et al., 2016). Além disso, em um trabalho anterior deste mesmo autor, os miRNAs miR-148a, miR-155 e miR-1321 foram sugeridos como potenciais alvos para terapias alternativas em PF (CIPOLLA e PETZL-ERLER, 2012).

Já os lncRNAs são transcritos longos (>200 nucleotídeos), também carentes de potencial codificador para proteínas, porém com uma vasta gama de mecanismos regulatórios, a nível transcricional e pós-transcricional (detalhes abordados no tópico seguinte). SNPs localizados em genes de lncRNAs foram associados ao PFE e preditos como moduladores da transcrição e estrutura secundária de seus lncRNAs, além de alterar sítios de ligação de miRNAs (LOBO-ALVES et al., 2019). Em outro estudo, um SNP do gene *LENG8* associado ao PFE foi predito de apresentar efeito regulatório *in cis* no gene de lncRNA antisenso *LENG8-AS1*, com possível co-regulação de expressão (FARIAS et al., 2018).

Em resumo, todos estes estudos demonstram a importância e complexidade dos fatores genéticos na etiologia do pênfigo. Contudo, poucos foram os que investigaram associações genéticas para ambos os pênfigos, foliáceo e vulgar. Considerando o complexo envolvimento dos lncRNAs na regulação da expressão gênica, assim como a existência de poucos estudos em PF e a ausência de estudos

em PV, ainda há muito para investigar a respeito destes transcritos, em ambas as doenças.

## 2.2 RNAS LONGOS NÃO CODIFICADORES

De acordo com o dogma central da biologia, a informação genética armazenada nos genes codificadores é transcrita em mRNAs, que por sua vez são traduzidos a proteínas, moléculas estas até então consideradas protagonistas nas funções e mecanismos celulares (CRICK, et al. 1961; CARNINCI, et al., 2005; revisado em SHI et al., 2013 e em WAPINSKI e CHANG, 2011). Dessa forma, acreditava-se que os mRNAs seriam a classe de RNAs mais abundante no transcriptoma. Contudo, com a intensa popularização dos sequenciamentos em larga escala e o crescente acesso a dados genômicos e de transcriptoma, tornou-se evidente que os RNAs não codificadores (ncRNAs) correspondem à maior parte dos transcritos em mamíferos (PONTING et al., 2009; GUTTMAN, et al., 2009; HARROW et al., 2012). Dentre as diversas classes de ncRNAs atualmente conhecidas, destacam-se os RNA ribossomais (rRNAs), RNAs transportadores (tRNAs), microRNAs (miRNAs), pequenos RNAs nucleares (snRNAs) e nucleolares (snoRNAs), pequenos RNAs de interferência (siRNAs), Piwi RNAs (piRNAs) e os RNAs longos não codificadores (lncRNAs) (PONTING et al., 2009).

Durante muito tempo, os ncRNAs foram considerados como ruído transcricional, e portanto, sem função biológica definida. Contudo, já está claro atualmente que estão envolvidos em diversos processos celulares e fisiológicos (MORAN et al., 2012). No presente trabalho, foram estudados os RNAs longos não codificadores (lncRNAs), uma das classes de ncRNAs mais abundantes e menos compreendidas do transcriptoma, que vem ganhando destaque na compreensão de diversas doenças (DERRIEN et al., 2012).

### 2.2.1 LncRNAs: definição e função

Os lncRNAs são definidos como transcritos não codificadores com mais de 200 nucleotídeos (KAPRANOV et al., 2007; HARROW et al., 2012). Os genes de lncRNAs diferenciam-se dos codificadores pela ausência de *open reading frames*

(ORFs) e, em geral, por apresentarem longos e poucos éxons. As regiões promotoras dos lncRNAs são bastante conservadas entre os vertebrados, diferentemente de seus éxons, que apresentam pouca conservação entre as mesmas espécies (PONTING et al., 2009; DERRIEN et al., 2012; HARROW et al., 2012; GENG e TAN, 2016). Essa baixa conservação não necessariamente significa baixa importância funcional: sugere-se que a alta variabilidade genética dos lncRNAs permita a origem de diferentes níveis de regulação e interação molecular dos mesmos (JOHNSSON et al., 2014; QUINN et al., 2016). Ainda assim, muitas variantes genéticas em lncRNAs vem sofrendo seleção purificadora, o que indica relevância funcional (PONJAVIC, PONTING & LUNTER, 2007). Além disso, diversos lncRNAs regulam a transcrição de genes vizinhos (*in cis*) durante o seu próprio processo de transcrição, de forma que em muitos casos, a localização do seu gene pode ser mais relevante a nível funcional, do que sua sequência nucleotídica propriamente dita (TOIBER, LEPRIVIER & ROTBLAT, 2017).

A localização do transcrito de lncRNA, que pode ser nuclear e/ou citoplasmática, está relacionada com a sua função na célula (regulatória transcricional/epigenética ou pós-transcricional). Muitos lncRNAs são transcritos pela RNA polimerase II, sofrem processamento alternativo, podem ou não ser poliadenilados, e podem apresentar expressão tecido- ou célula-específica, indicando que são altamente regulados (DERRIEN et al., 2012; HARROW et al., 2012; SALVIANO-SILVA & LOBO-ALVES, 2018). Sua estrutura secundária é formada através de dobras moleculares, resultantes do pareamento de bases entre sequências complementares, que dinamicamente expõem motivos funcionais que permitem a interação do lncRNA com outras moléculas (SMITH & MATTICK, 2016; PEGUEROLES & GABALDÓN, 2016).

De acordo com o catálogo de lncRNA do NONCODE (FANG et al., 2017; disponível em http://www.noncode.org/analysis.php, acessado em 02/2020), já são reportados mais de 96.000 genes de lncRNA e mais de 172.000 transcritos de lncRNAs em humanos. Estes números podem ser ainda maiores, considerando que lncRNAs também podem ser transcritos a partir de alguns pseudogenes, e que certos lncRNAs são processados a pequenos ncRNAs (DERRIEN et al., 2012; WANG et al., 2011; MILLIGAN & LIPOVICH 2015).

De acordo com a sua localização em relação ao gene codificador de proteína mais próximo, os genes de lncRNAs podem ser categorizados como: intergênicos (lincRNAs), quando situados em *loci* vizinhos aos de genes codificadores, sem sobreposição; antisenso, quando localizados sobrepostos na fita oposta ao gene codificador de proteína, podendo ser intrônicos (completamente dentro dos limites do intron da fita oposta, sem sobreposição com éxons) ou *natural antisense transcripts* (NAT - quando se sobrepõem ao éxon da fita oposta, em pelo menos 1pb); sobrepostos senso, quando localizados na mesma fita e transcritos na mesma direção do gene codificador de proteína sobreposto, podendo ser intrônicos ou de sobreposição com éxons; ou bidirecionais (divergentes), quando localizados na fita oposta ao gene codificador de proteína, cujos sítios de início de transcrição (TSS) são próximos (SALVIANO-SILVA & LOBO-ALVES et al., 2018; HRDLICKOVA et al., 2014) (FIGURA 4A).

Um grande número de lncRNAs possuem padrões específicos de expressão, indicando que são altamente regulados e envolvidos em funções biológicas específicas, apesar de pouco conservados (SALVIANO-SILVA & LOBO-ALVES et al., 2018; DINGER et al., 2008; MERCER et al., 2008). Em geral, os lncRNAs estão envolvidos em diversos processos regulatórios (FIGURA 4B), onde atuam por diferentes mecanismos funcionais, modulando a transcrição gênica, processamento, tradução, atividade e transporte de moléculas, *imprinting*, entre outros processos celulares. Os lncRNAs podem ainda ser processados em pequenos RNAs, como os miRNAs. Além disso, os lncRNAs também estão envolvidos em mecanismos epigenéticos, induzindo a remodelagem da cromatina e alterando o recrutamento de proteínas cruciais para a expressão gênica (SALVIANO-SILVA & LOBO-ALVES et al., 2018; SHI et al., 2013; WAPINSKI & CHANG, 2011).

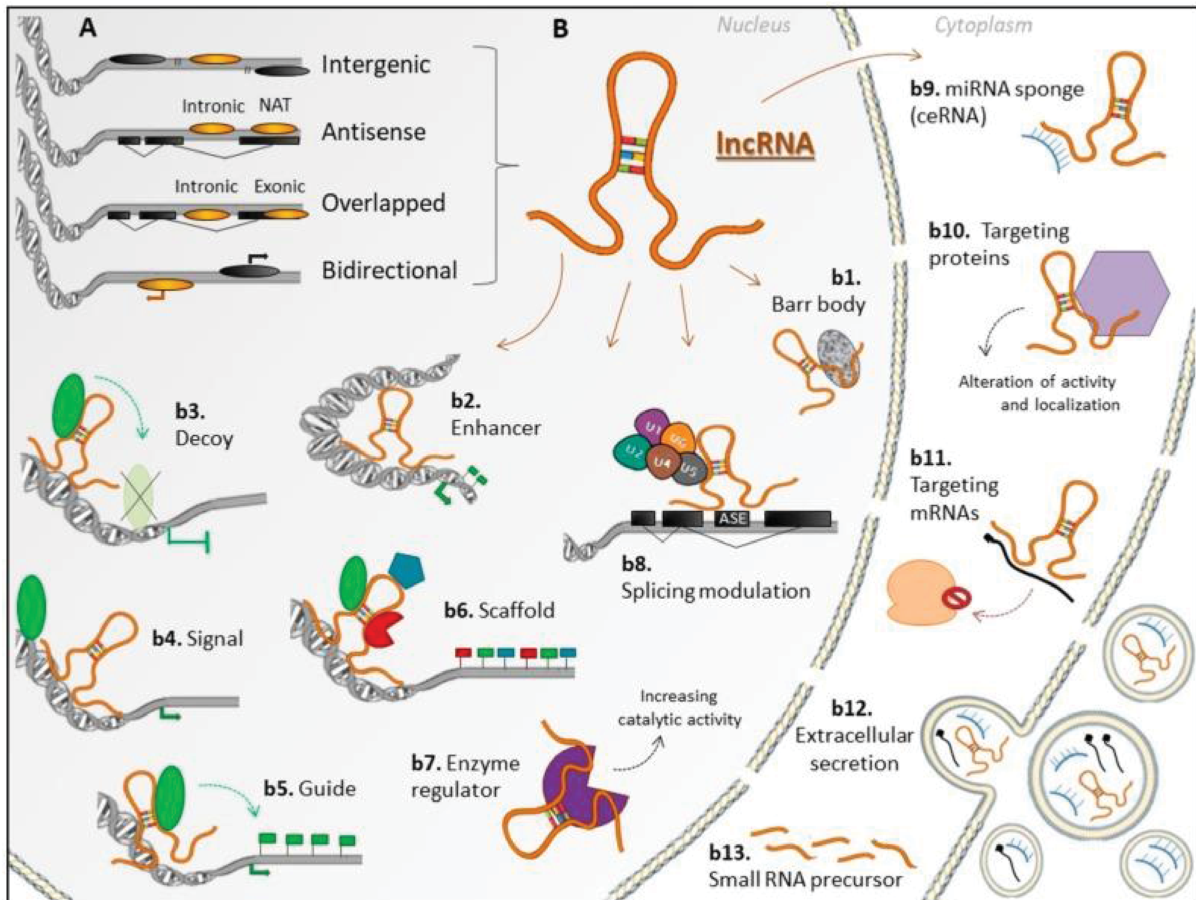



FIGURA 4. Classificação e mecanismos regulatórios dos lncRNAs na célula. **(A).** Classificação dos genes de lncRNAs (elipses amarelas) de acordo com a localização genômica do gene codificador de proteína mais próximo (elipses pretas) e/ou do exon do gene codificador. **(B)** Principais funções e mecanismos regulatórios exercidos por lncRNAs: (b1) atuação no processo de inativação do cromossomo X adicional em mulheres; (b2) atuação como acentuadores (*enhancers*), induzindo a transcrição de um gene *in cis* ou *in trans*; (b3) sequestrando proteínas regulatórias (como fatores de transcrição e modificadores de cromatina), impedindo sua ligação ao DNA; (b4) sinalizando para vias regulatórias, e assim ativando ou inativando genes; (b5) guiando proteínas regulatórias para sítios específicos; (b6) ligando-se a diferentes proteínas e formando complexos ribonucleoproteicos, que também poderão afetar a expressão gênica; (b7) interagindo com enzimas, como as quinases, influenciando em seu poder catalítico e/ou alterando sua sinalização; (b8) modulando o processamento (*splicing*) alternativo de transcritos primários; (b9) atuando como ceRNAs (*competing endogenous RNA*) através do sequestro de miRNAs, impedindo suas funções regulatórias; (b10) ligando-se a proteínas , bloqueando ou induzindo suas funções, ou ainda alterando sua localização na célula; (b11) ligando-se a mRNAs, estabilizando-o, degradando-o ou ainda impedindo sua tradução no ribossomo. Além disso, os lncRNAs podem ainda (b12) ser transportados através de vesículas extracelulares (VEs) para outras células e tecidos, onde poderão realizar seus mecanismos regulatórios; (b13) ser precursores de pequenos RNAs regulatórios, como miRNAs. Um lncRNA pode apresentar múltiplos mecanismos regulatórios, no núcleo e/ou citoplasma. Em b12, o transporte através de VEs não é por si só um mecanismo regulatório, mas possibilita que o lncRNA atue em processos regulatórios em

tecidos em que o mesmo não estaria sendo expresso. ASE—alternatively spliced exon. Fonte: SALVIANO-SILVA & LOBO-ALVES et al. (2018).

### 2.2.2 LncRNAs e doenças

Como já mencionado previamente, ainda é pouco o que se conhece sobre os mecanismos dos lncRNAs, contudo, sabe-se que participam de uma série de processos biológicos e regulatórios. Sabe-se ainda que a desregulação de alguns lncRNAs pode ser também uma consequência destes processos, apresentando-se como importantes potenciais biomarcadores de falhas regulatórias (GUTTMAN, et al., 2009; WILUSZ et al., 2009).

Em humanos, mutações na sequência primária dos lncRNAs, bem como alterações na sua expressão, têm sido associadas com diversas eventos patológicos, com destaque para síndromes, distúrbios neurodegenerativos, cardiovasculares, autoimunes e inflamatórios (ESTELLER, 2011; HRDLICKOVA et al., 2014; MORAN et al., 2012; RICAÑO-PONCE e WIJMENGA, 2013; CIPOLLA et al., 2018), e principalmente em vários tipos de câncer (OLIVEIRA et al., 2019). As numerosas informações que vem sendo reportadas sobre a desregulação de lncRNAs no câncer, demonstram que muitos destes transcritos exercem importantes função oncogênicas e oncossupressoras (GIBB et al., 2011; OLIVEIRA et al., 2019).

Dentre as diversas doenças não cancerosas relacionadas com alterações nos lncRNAs, as doenças autoimunes vem ganhando um crescente destaque. Estudos de associação do genoma total (GWAS) têm relacionado diversos SNPs com doenças imunológicas (MAURANO et al., 2012), dos quais, cerca de 10% destes SNPs associados encontram-se em *loci* de lncRNAs, o que sugere que estes transcritos exercem um papel na etiologia destas doenças (RICAÑO-PONCE e WIJMENGA, 2013). Contudo, apesar da lista de lncRNAs relacionados a doenças autoimunes estar em considerável expansão, a compreensão detalhada de seus mecanismos na patogênese destas desordens, ainda é bastante limitada.

#### *2.2.2.1 lncRNAs em doenças autoimunes*

Vários lncRNAs participam de processos fisiológicos e ontológicos em diferentes células do sistema imune, indicando um importante papel dos lncRNAs no desenvolvimento e homeostase deste sistema (SALVIANO-SILVA & LOBO-ALVES et al., 2018). Alterações genéticas e de expressão destes lncRNAs poderiam, portanto, interferir em diferentes processos de interação molecular, levando assim a uma desregulação imunológica. De fato, um crescente número de lncRNAs contribuem com a susceptibilidade ou proteção a diversas doenças autoimunes (CIPOLLA et al., 2018).

Foi estimado que aproximadamente 7% dos SNPs associados com doenças autoimunes parecem relacionados com lncRNAs intergênicos (lincRNAs), sendo que muitos destes SNPs apresentam efeitos na expressão (eQTL, do inglês *Expression. Quantitative Trait Loci*) destes lincRNAs (KUMAR et al., 2012; RICAÑO-PONCE e WIJMENGA, 2013), podendo ainda influenciar a expressão de proteínas envolvidas nestas doenças (HRDLICKOVA et al., 2014). De fato, diversas variantes associadas com doença celíaca, psoríase, doença de Chron, esclerose múltipla, artrite reumatoide, lúpus eritematoso sistêmico (SLE), colite ulcerativa, diabetes tipo 1, entre outras, ocorrem em loci de lincRNAs (inclusive em regiões eQTLs), como já evidenciado em estudos de GWAS (RICAÑO-PONCE & WIJMENGA, 2013).

Alguns dos SNPs mais fortemente associados com a doença celíaca encontram-se em genes de lincRNAs, miRNAs e snoRNAs (KUMAR et al., 2012). Já os SNPs fortemente associados com a aterosclerose encontram-se próximos e correlacionados com o lncRNA *ANRIL* (*Antisense Non-coding RNA in the INK4 Locus*), e parecem regular o risco da aterosclerose por modular a expressão e/ou estrutura do *ANRIL* (BIRD et al., 2010).

A maior expressão do lncRNA *GAS5* (*Growth Arrest-Specific 5*) em células do sistema imune diminui a atividade transcricional induzida por fatores de transcrição de receptores de glicocorticoides, diminuindo a imunossupressão e contribuindo para o desenvolvimento da autoimunidade. O *locus* do *GAS5* tem sido, portanto, associado com a susceptibilidade a doenças autoimunes, como o SLE (KINO et al., 2010). Outros lncRNAs associados ao SLE foram encontrados, apresentando-se com expressão desregulada e correlacionados com genes próximos também desregulados, em comparação com indivíduos saudáveis (SHI et al., 2014).

O lncRNA regulatório *PRINS* (*Psoriasis associated non-protein coding RNA Induced by Stress*) é regulado pela proliferação e diferenciação dos queratinócitos e possui um papel protetor em células expostas ao estresse. Níveis elevados de *PRINS* foram associados com a presença de epiderme lesionada de pacientes com psoríase (SONKOLY et al., 2005). Já na artrite reumatoide, foi reportado um aumento de expressão do lncRNA *H19* (*imprinted maternally expressed transcript*) no tecido sinovial (STUHLMULLER et al., 2003) e do lncRNA *HOTAIR* nas células mononucleares sanguíneas e nos exossomos séricos de pacientes. Em contraste, a expressão de *HOTAIR* (*HOX Transcript Antisense RNA*) encontra-se diminuída em osteoclastos diferenciados e sinoviócitos. Estes dados evidenciam que *HOTAIR* é um potencial biomarcador no diagnóstico da artrite reumatoide (SONG et al., 2013).

Diversos lncRNAs tem sido relacionados com a regulação, diferenciação e respostas de células T $CD4^+$, em nível fisiológico e patológico (WEST & LAGOS, 2019). Dentre estes lncRNAs, o IFNG-AS1 (*Interferon-Gamma Antisense transcript 1*) regula a expressão de IFN-$\gamma$ em células Th1 (WEST & LAGOS, 2019). Além disso, maiores níveis de IFNG-AS1 podem diminuir a expressão de *HLA-DRB* e *HLA-DOB* em miastenia grave, diminuindo assim a expressão de *CD40L* e regulando a ativação de células T $CD4^+$ (LUO et al., 2017).

Além destes e de tanto outros exemplos disponíveis na literatura, foi observado ainda que em células do sistema imune, há uma maior expressão de lncRNAs localizados em *loci* associados a nove distúrbios imunes, em relação a lncRNAs localizados em outras regiões do genoma (HRDLICKOVA et al., 2014b). Dessa forma, torna-se evidente a relevância da compreensão dos mecanismos regulatórios dos lncRNAs, cuja correlação e interação com outras moléculas mostra-se promissora na predição de vias de sinalização de distúrbios imunes e na busca por potenciais alvos terapêuticos.

#### 2.2.2.2 *lncRNAs em pênfigo*

Como previamente comentado, a investigação de lncRNAs é bastante recente em PF, e ainda ausente em PV. Em PF, há apenas dois estudos com lncRNAs publicados até o momento, sendo estes limitados a associação com polimorfismos genéticos, não explorando ainda a expressão destes transcritos.

No primeiro, 2080 SNPs de lncRNAs foram analisados em um estudo de associação com 229 pacientes de PFE e 6681 controles. O SNP rs7144332**T* localizado no gene de lncRNA *AL110292.1* foi associado com maior susceptibilidade à doença ($P < 2.4 \times 10^{-5}$, OR = 1.63), enquanto SNPs dos lncRNAs *lnc-PREX1-7:1*, *AC009121.1*, *AC133785.1*, *LINC01176* e *LINC01119* foram considerados como sugestivos de associação ($0.001 > P > 2.4 \times 10^{-5}$) (LOBO-ALVES et al., 2019). No segundo trabalho sobre lncRNAs em PFE, realizado pela mesma autora, 4 SNPs localizados na região genômica *1p13.2* foram analisados em 230 pacientes de PFE e 190 controles. Destes, dois SNPs estão localizados no lncRNA *AL137856.1* e um SNP no lncRNA *AP4B1-AS1* (*Adaptor related Protein complex 4 subunit beta 1 Antisense transcript 1*), sendo todos sobrepostos com genes codificadores de proteínas. Nenhuma associação genética foi encontrada para os SNPs analisados (LOBO-ALVES et al., 2019 [2]).

Além disso, em um estudo de rastreamento da região genômica do complexo de receptores leucocitários (LCR) em PFE, foi predito que o SNP associado *rs35336528*G* , localizado no gene codificador *LENG8*, apresenta efeito eQTL na expressão do lncRNA antisenso *LENG8-AS1*, com possível efeito regulador *in cis* na expressão de ambos os genes (FARIAS et al., 2018). Contudo, uma validação desta predição ainda não foi investigada.

## 2.3 GENES *HLA COMPLEX GROUP* (HCG)

O MHC é uma região de aproximadamente 4Mb, presente no braço curto do cromossomo 6 (especificamente localizado em *6p21.3*), onde estão localizados os genes HLA e outros genes de grande importância em processos imunológicos, sendo a maioria destes genes altamente polialélicos e polimórficos. Por este motivo, esta região é intensamente investigada em diversas doenças autoimunes, inflamatórias e infecciosas, além da importância clínica de genes HLA para transplantes (TRON et al., 2005; KULSI, SHIINA & DIJKSTRA, 2019). O MHC é

dividido em três regiões genômicas: o MHC de classe I, onde estão localizados genes de HLA classe I clássicos (*HLA-A, -B* e *–C*) e não clássicos; MHC de classe II, localizado em região mais próxima ao centrômero, e que abriga os genes HLA classe II (*HLA-DR, HLA-DQ, HLA-DP*, e outros); e por fim, intermediário ao MHC I e MHC II, localiza-se o MHC de classe III, que apesar de não possuir genes HLA clássicos, abriga genes conhecidamente importantes no contexto imunológico, como o grupo gênico do fator de necrose tumoral (*TNF*), genes dos componentes *C2* e *C4* do sistema complemento, dentre outros (GOLDBERG & RIZZO, 2015).

Além dos aqui citados, há ainda uma ampla variedade de genes codificadores e não codificadores nestas regiões (KULSKI, SHIINA & DIJKSTRA, 2019). Dentre os não codificadores, o lncRNA *HCP5* (*HLA complex pseudogene 5*) já foi associado com algumas doenças autoimunes infecciosas (KULSKI, 2019), enquanto o lncRNA *PSORS1C3* (*psoriasis susceptibility 1 candidate 3*) está associado com a susceptibilidade à psoríase (HOLM et al., 2005). Ambos os genes também são sugeridos de conferir susceptibilidade diferencial à artrite reumatoide associada à esquizofrenia (MALAVIA et al., 2017).

Na história da investigação de genes não clássicos na região do MHC, alguns dos primeiros estudos envolvendo genes hoje descritos como lncRNAs, se iniciaram nos anos 90, durante a caracterização da família gênica do MIC (*MHC class I polypeptide-related sequence*, também conhecido como *PERB11*). Nestes trabalhos, genes presentes no agrupamento do MHC I foram descobertos. Por fazerem parte de uma região altamente associada com hemocromatose, foram denominados "*hemochromatosis candidate gene*" (HCG), e enumerados de 1 a 7 (HCG I-VII) (KAHLOUN et al., 1993). A partir de então, novos genes HCG foram descritos, como o HCG IX (*HCG9*), que está relacionados a membros do cluster MIC (PICHON et al., 1996; PICHON et al., 1996 [2]). Anos depois, a sigla HCG teve o seu significado alterado, sendo agora classificada como *HLA Complex Group* (HCG) (HORTON et al., 2004).

Atualmente, genes do HCG já foram descritos em uma ampla extensão genômica, abrangendo desde a região a jusante do MHC I (*6p21.1*) até o MHC II. A família multigênica dos HCGs é composta por genes codificadores de proteínas (com nomes diversos), pseudogenes (como o *HCG4* e *HCG4B*) e genes de lncRNAs

(*HCG9, 11, 14, 15, 17, 18, 20, 21, 22, 23, 24, 25, 26 e 27*) (RNA Central DB - THE RNACENTRAL CONSORTIUM, 2019; Ensembl Genome Browser - HOWE et al., 2019). A disposição genômica dos lncRNAs HCGs em *6p21.1 - 21.3* está representada na FIGURA 5.

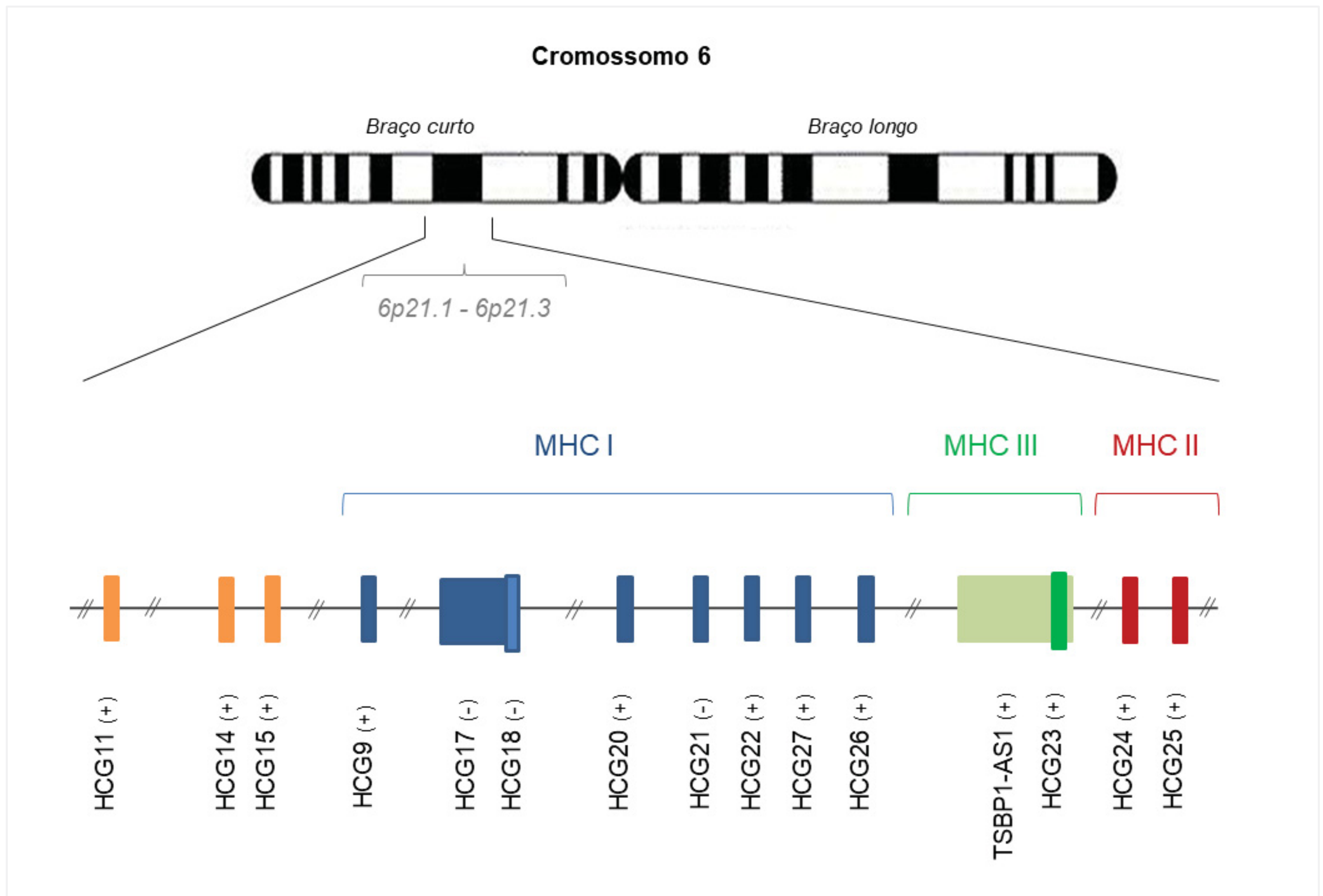



FIGURA 5. Localização genômica dos genes HCG no braço curto do cromossomo 6. Em laranja, estão representados os genes HCG localizados na região 6p.21.1-21.2, a jusante da região MHC (com início em 6p21.3); em azul, estão representados os genes HCG localizados na região do MHC classe I; em verde, o gene HCG23 presente na região do MHC classe III, sobreposto ao gene de lncRNA TSBP1-AS1; em vermelho, genes HCG localizados no MHC classe II. Demais genes sobrepostos na região não estão sendo representados na figura. As localizações genômicas são meramente ilustrativas, não representando a escala real da região. Os símbolos "(+)" e "(-)" indicam a orientação da fita onde o gene está inserido (direta ou reversa, respectivamente). Fonte: a autora (2020).

Devido à grande quantidade de informações geradas sobre algumas regiões genômicas, diferentes transcritos de RNA vêm sendo atualizados nos diferentes

bancos de dados, atualizações estas que nem sempre coincidem entre os mesmos, podendo causar confusão. Este é o caso do gene *HCG23*, um lncRNA antisenso ao gene codificador de proteína *C6orf10*, e que apresenta informações contrastantes nos principais bancos de dados genômicos.

De acordo com a versão genômica hg19 do Ensembl, *HCG23* (ENSG00000228962) é um pequeno gene responsável pela expressão do lncRNA HCG23-001 (ENST00000426643.1) de 610pb [1]. Já na versão genômica atual (hg38) deste site, ENST00000426643.1 (agora classificado como TSBP1-AS1-204) é um dos 15 transcritos originados a partir do gene de lncRNA *TSBP1-AS1* (agora ENSG00000225914), antisenso a uma grande parte dos genes *C6orf10* (agora também denominado TSBP1 - *testis-expressed basic protein 1*) e *BTNL2* (*butyrophilin-like protein 2*) [2]. Estas informações se confirmam pelo banco de dados RNA Central (hg38), onde estão representados os diferentes transcritos de *TSBP1-AS1*, incluindo o até então HCG23 (TSBP1-AS1-204) [3]. Já de acordo com o LNCipedia, o gene *TSBP1-AS1* não é encontrando, sendo os seus transcritos derivados do gene *HCG23*, o qual também possui os códigos ENSG00000228962 (hg19) e ENSG00000225914 (hg38) [4]. Já no banco de dados GTEX (hg38), informações de expressão constam tanto para ENSG00000228962 quanto para ENSG00000225914: o primeiro é encontrado como *HCG23* (transcrito ENST00000426643.1), e o segundo gene também como *HCG23*, além de sua aliase *XXbac-BPG154L12.4*, com informação apenas para o transcrito ENST00000425033.1 (HCG23:19 ou TSBP1-AS1-203, de acordo com LNCipedia ou Ensembl, respectivamente) (GTEX Portal). Por fim, nesta mesma região genômica, consta ainda o gene *LOC101929163*, classificado como HCG23:6 [5], cujo transcrito possui 807pb. Contudo, não há qualquer informação sobre este RNA em ambas as versões genômicas do Ensembl, apesar do mesmo ser citado em diversos estudos.

Dessa forma, não está claro se *HCG23*, *LOC101929163* (HCG23:6) e *TSBP1-AS1* são genes distintos, apesar de ocuparem, em parte, a mesma posição genômica e compartilharem seus éxons (FIGURA 6); ou se HCG23 e LOC101929163 são atualmente considerados transcritos do gene *TSBP1-AS1*, como produtos de processamento alternativo.

1 disponível em http://grch37.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000228962;r=6:32358287-32361463;t=ENST00000426643
2 disponível em http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000225914;r=6:32254640-32407763
3 disponível em https://rnacentral.org/rna/URS000075B48F/9606
4 disponível em https://hg19.lncipedia.org/db/search?search_id=hcg23
5 disponível em https://rnacentral.org/rna/URS0000A764AC/9606 e em https://hg19.lncipedia.org/db/search?search_id=hcg23%3A6)

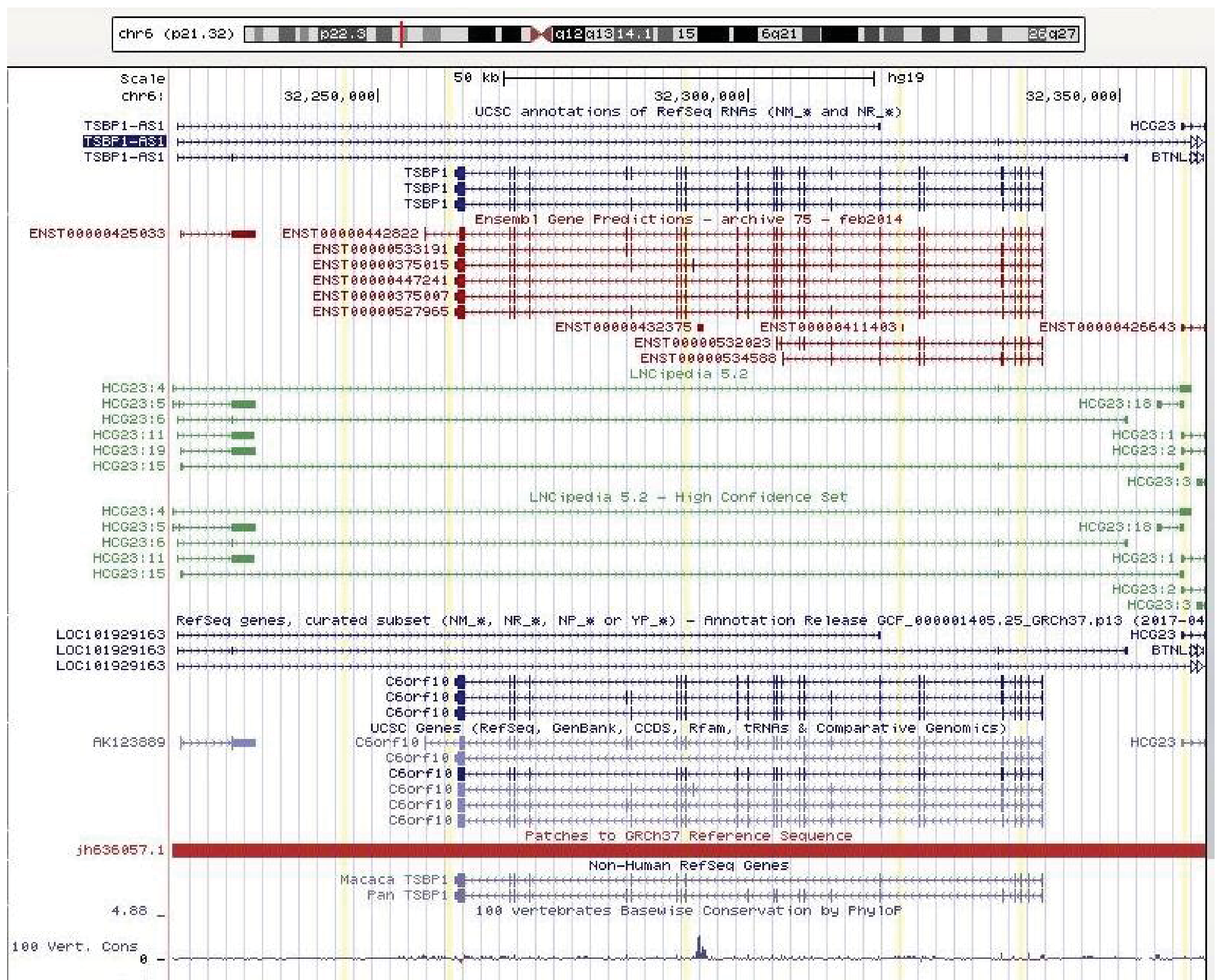



FIGURA 6. Região genômica do ENSG00000225914 (hg19 - chr6:32,222,417–32,361,468). De acordo com os diferentes bancos de dados, lncRNAs classificados com diferentes nomes são expressos a partir da mesma região. De acordo com UCSC/RefSeq (em azul): TSBP1-AS1 como os principais transcritos (acima); o pequeno transcrito de HCG23 (na extremidade direita); LOC101929163 (centro); e os codificadores C6orf10, também nomeados como TSBP1. De acordo com Ensembl (em vermelho): transcritos de TSBP1/C6orf10 (região central) e de HCG23 (ENST00000426643, à direita). De acordo com LNCipedia (em verde): transcritos de HCG23 expressos a partir de toda a extensão genômica. Fonte: UCSC genome browser (Disponível em: https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr6%3A32222417%2D32361468&hgsid=797904707_Qum8WGVp3oHQKfNR3xr1gwm9SRzq).

Apesar da divergência de informações, a maior parte dos estudos que abordam o gene *TSBP1-AS1* (e seus diferentes transcritos) se refere ao mesmo como *HCG23* e/ou *LOC101929163*. Dentre estes trabalhos, variantes genéticas nesta região já foram associadas com susceptibilidade diferencial à artrite reumatoide (ANAPARTI et

al., 2019), psoríase (FENG et al., 2009), colite ulcerativa (MOON et al., 2018), síndrome nefrótica (DEBIEC et al., 2018), doença de Graves (KHONG et al., 2016), esclerose múltipla (ZILIOTTO et al., 2019), lúpus neonatal (SAXENA et al., 2012; CLANCY et al., 2010), e com a idade de manifestação da expansão de *C9orf72* (causal em esclerose lateral amiotrófica e demência frontotemporal) (ZHANG et al., 2018).

Com relação ao lncRNA *HCG22*, SNPs presentes neste gene foram associados com esclerose múltipla (LIN et al., 2015), lúpus (CHUNG et al. 2014) e cardiomiopatia dilatada (MEDER et al., 2013). Além destes, SNPs próximos à região genômica do *HCG22* também foram associados com asma de início tardio (YATAGAI et al., 2016) e com a instalação e progressão do HIV em indivíduos do sul da África (XIE et al., 2017; THAMI et al., 2019).

Em diabetes mellitus tipo 2, um SNP associado com a doença foi descoberto por estar próximo ao gene *HCG27*, que foi então avaliado de acordo com sua expressão. Os níveis transcricionais de *HCG27* apresentaram-se diminuídos em PBMCs de pacientes (em comparação com controles), e negativamente correlacionados com níveis glicêmicos. Além disso, análises de potencial diagnóstico sugerem *HCG27* como um possível biomarcador para esta doença (SAEIDI et al., 2018). Outros SNPs próximos a *HCG27*, presentes na região do gene *PSORS1*, foram associados com psoríase. Além disso, a região do *HCG27* parece apresentar um padrão de metilação característico de um perfil epigenético relevante para a ativação de células T (CLOP et al., 2013). Já o gene *HCG26* encontra-se hipometilado em artrite psoriática, em relação a indivíduos com psoríase e controles. Alguns transcritos de *HCG26* possuem ainda sítios de inserção para elementos transponíveis, o que também acarreta em diferenças em seus padrões de metilação (POLLOCK et al., 2019).

Ainda em psoríase, um SNP próximo à região gênica de *HCG9*/*MICA*/*MICB* foi associado com a susceptibilidade a esta doença (KNIGHT et al., 2012). SNPs presentes na região do *HCG9* e *HLA-A* foram ainda associados com linfoma de Hodgkin Epstein-Barr positivos (NIENS et al., 2006; URAYAMA et al., 2012), entre outros tipos de câncer.

Em doença celíaca, SNPs do gene *HCG14* foram associados com esta enfermidade e podem estar correlacionados com a expressão diferencial de *NOD1* (*Nucleotide-binding Oligomerization Domain-containing protein 1*). O silenciamento de *HCG14* levou a uma ligeira diminuição nos níveis de mRNA de NOD1. Devido à localização preferencialmente nuclear do HCG14, os autores sugerem uma possível função de regulação a nível transcricional por parte deste lncRNA (SANTIN et al., 2018).

Juntamente a outros genes de lncRNA, o *HCG17* foi sugerido como um potencial biomarcador para acidente vascular cerebral isquêmico, no qual o mesmo encontra-se menos expresso (ZHENG et al., 2019). Por fim, estudos recentes tem relacionado o HCG11 como competidor endógeno (ceRNA) para diferentes miRNAs em alguns tipos de câncer, como glioma (CHEN et al., 2019; ZHANG, L. et al., 2019), câncer de próstata (ZHANG, H. et al, 2019), entre outros. Contudo, o envolvimento deste e de outros genes HCG em doenças imuno-relacionadas ainda está para ser investigado.

Em suma, apesar da região do MHC ser bastante reconhecida por sua importância em diversos processos imunológicos, ainda há uma carência de estudos explorando a função dos genes HCG. Apesar disso, muitos estudos de associação (especialmente GWAS) têm revelado associações genéticas entre estes genes e diferentes tipos de doenças, demonstrando uma potencial função dos HCGs em processos fisiológicos e patológicos do sistema imune.

## 3. HIPÓTESE E JUSTIFICATIVA

O pênfigo é um grupo de doenças bolhosas graves, cuja etiologia é bastante complexa e apenas parcialmente conhecida. Apesar de negligenciado, o PFE é a única doença autoimune conhecida por ser endêmica, indicando a presença de potentes fatores ambientais em certas nas regiões, que por sua vez podem desencadear a doença em indivíduos geneticamente susceptíveis. Além disso, o PFE parece apresentar diferenças de susceptibilidade genética em relação à sua forma esporádica (e ambas as formas de PF também possuem diferenças de susceptibilidade ao PV). O conhecimento atual sobre os fatores genéticos associados ao pênfigo é praticamente restrito a genes codificadores de proteínas, dentre os quais se destacam os genes HLA de classe II, presentes na importante região do MHC. Contudo, além dos genes codificadores, os genes de lncRNAs vem ganhando uma crescente atenção na compreensão de diversos mecanismos moleculares, por apresentarem diferentes funções regulatórias transcricionais e pós-transcricionais. Variantes genéticas e níveis alterados de expressão de lncRNAs tem sido cada vez mais investigados em eventos fisiológicos e patológicos, inclusive na autoimunidade. Além disso, diversos genes de lncRNAs estão presentes na região genômica do MHC, dentre os quais, os lncRNAs classificados como *HLA-Complex Group* (HCG) incluem alguns genes já associados com doenças autoimunes, apesar de ainda pouco conhecidos. Devido à importância da região genômica do MHC em processos imunológicos e autoimunes, assim como à natureza regulatória dos lncRNAs, é possível que os genes HCGs desempenhem um papel importante na autoimunidade do pênfigo. Considerando esta hipótese, se faz justificável a investigação genética e de expressão dos genes HCGs em pênfigo foliáceo e vulgar. Destaca-se que este estudo é pioneiro em dois aspectos: na investigação de todos genes de lncRNAs HCGs atualmente descritos e na investigação de lncRNAs em PV. Os resultados obtidos neste estudo serão importantes pois se destinam a explorar uma região ainda pouco investigada dentro do MHC, tendo o potencial de indicar novos genes candidatos para estudos de doenças autoimunes e imuno-relacionadas e permitirão ampliar o conhecimento acerca dos fatores genéticos envolvidos no pênfigo.

# 4. OBJETIVOS

## 4.1 OBJETIVO GERAL

O objetivo geral dessa tese de doutorado é de investigar se a variabilidade genética dos lncRNAs HCG, bem como se os níveis de expressão dos HCGs associados, influenciam a susceptibilidade aos pênfigos foliáceo e vulgar.

## 4.2 OBJETIVOS ESPECÍFICOS:

**I** – Selecionar variantes localizadas nos genes de HCGs a partir de estudos de genoma total e investigar se estas estão associadas com PF e PV através de estudo caso-controle;

**II** – Se encontrados genes associados, avaliar se estão diferencialmente expressos em amostras de pacientes de PF e PV, em relação a indivíduos controle;

**III** – Avaliar se genes HCG associados estão diferencialmente expressos em queratinócitos funcionalmente representitativos de PF e PV, após serem estimulados em cultivo com anticorpos de pacientes e controles.

# 5. CAPÍTULO I:

VARIABILIDADE GENÉTICA E EXPRESSÃO DOS RNAS LONGOS NÃO CODIFICADORES *HLA COMPLEX GROUP* (HCGS) NOS PÊNFIGOS FOLIÁCEO E VULGAR.

ORIGINAL ARTICLE

# HLA-COMPLEX GROUP (HCG) LNCRNA GENES: GENETIC ASSOCIATIONS AND DIFFERENTIAL EXPRESSIONS IN PEMPHIGUS DISEASES

Amanda Salviano-Silva[1], Mareike Becker[2], Danillo G. Augusto[1,#], Hauke Busch[2,##], Matthias Munz[2], Rodrigo Coutinho de Almeida[1,###], Maria Luiza Petzl-Erler[1], Nina Van Beek[3], Detlef Zillikens[3], Enno Schmidt[2,3], Angelica B. W. Boldt[1], Saleh Ibrahim[2], Danielle Malheiros[1].

[1] *Postgraduate Program in Genetics. Department of Genetics, Federal University of Paraná (UFPR), Curitiba, Brazil.*

[2] *Lübeck Institute for Experimental Dermatology, Lübeck, Germany.*

[3] *Department of Dermatology, University of Lübeck, Lübeck, Germany.*

Current address:

[#] *Department of Neurology, University of California San Francisco, USA.*

[##] *Institute for Cardiogenetics, University of Lübeck, Lübeck, Germany.*

[###] *Department of Biomedical Data Sciences, Sector Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands.*

**Corresponding author**

Dr. Danielle Malheiros

Departamento de Genética

Universidade Federal do Paraná

Caixa Postal 19071

81531-980 Curitiba, Brasil

Tel: (+5541) 3361-1553, 3361-1724

E-mail: dani_malheiros@ufpr.br

**ABSTRACT**

**Background:** Pemphigus is a group of autoimmune diseases characterized by the autorecognition of desmosomal antigens, such as desmoglein-1 (DSG1, for pemphigus foliaceus - PF) and -3 (DSG3, for pemphigus vulgaris - PV), leading to acantholysis and skin blisters. Several studies have shown associations of genetic variants (alleles, haplotypes, or individual single nucleotide polymorphisms - SNPs) with pemphigus foliaceus (PF) and pemphigus vulgaris (PV) susceptibility, especially alleles of class II human leukocyte antigen (HLA) genes and other major histocompatibility complex (MHC) genes. Recently, SNPs located in long non-coding RNAs (lncRNAs) have also been associated with many autoimmune diseases, including PF. However, none of these studies included a group of lncRNA genes located in the MHC region: the HLA complex group (HCG) genes. **Objectives:** We investigated for the first time if SNPs in HCG lncRNA genes, as well as HCG lncRNA levels, are associated with the susceptibility for PF and PV. **Methods and results:** We analyzed SNPs located in 13 HCG lncRNA genes (genotyped by Illumina microarrays), both in endemic PF (EPF) (227 patients and 194 controls from Brazil) and PV (241 patients and 1188 controls from Germany), applying multivariate logistic regression. We found 9 SNPs associated with EPF (FDR $p<0.05$): 7 located in the *TSBP1-AS1* lncRNA gene (which hosts *HCG23*); one in *HCG27* and one in *HCG21*. Moreover, 55 SNPs were associated with PV (FDR $p<0.01$), overlapping the genes *TSBP1-AS1* (38 SNPs), *HCG17* (6), *HCG27* (5), *HCG22* (3), *HCG20* (2) and *HCG26* (1). We therefore evaluated the *TSBP1-AS1*, *TSBP1*, *HCG23* and *HCG27* RNA levels in peripheral blood mononuclear cells (PBMC) of EPF and PV patients from Brazil, in comparison with controls. We observed *HCG27* upregulated in EPF ($p=0.035$, log2 FC = 1.3), while *TSBP1-AS1* was downregulated in PV ($p=0.029$, log2 FC = -1.29). The same differences were also seen in cultured immortalized keratinocytes stimulated with serum IgG antibodies from patients and controls (5 sporadic PF, 5 PV and 5 controls from Germany). Meanwhile, TSBP1 mRNA levels were decreased in EPF PBMCs ($p=0.042$, log2 FC = -2.14), although no expression differences were found for this gene in sporadic PF (keratinocytes). **Conclusions:** HCG genes are associated with the genetic susceptibility for both PF and PV, being *HCG27* and *TSBP1-AS1* also differentially expressed in the respective diseases, in different population samples. These results suggest a role of HCG lncRNAs in pemphigus autoimmunity.

Keywords: HLA Complex Group, lncRNAs, pemphigus, genetic susceptibility.

# 1. INTRODUCTION

Pemphigus is a group of serious epidermal diseases and one of the four main groups of autoimmune bullous diseases (AIBD) (Diaz et al., 1989; Sinha, 2011). Pemphigus is characterized by the presence of epidermal autoantibodies that recognize desmossomal antigens in the surface of keratinocytes, which are important to intercellular communication. The pathogenic auto-recognizing of these adhesion molecules results in detachment between keratinocytes, a process known as acantholysis, with clinical manifestation of epithelial blistering (Takeichi, 1991; Culton et al., 2008; Nisihara et al., 2003; Kasperkiewicz et al., 2017; Pollmann et al., 2018; Spindler et al., 2018; Schmidt, Kasperkiewicz & Joly, 2019). There are two classical pemphigus types: pemphigus foliaceus (PF) and pemphigus vulgaris (PV). In PF, the main autoantigen is Desmoglein 1 (Dsg1), highly expressed in the epithelial superficial layer, causing superficial skin blistering lesions, rare in the mucosa. In the case of PV, the pathogenic autoantibodies mainly recognize Desmoglein 3 (Dsg3), which is expressed in deeper epithelial layers. This results in profound epidermal blistering, affecting also mucous membranes (Diaz et al., 1989; Bystryn & Rudolph, 2005; Schmidt, Kasperkiewicz & Joly, 2019).

Regarding epidemiological aspects, pemphigus diseases have a global incidence of 0.75-5 cases/million per year. Nevertheless, besides rarer than PV throughout the world, PF is endemic in certain rural areas of Brazil, Colombia and Tunisia. In Brazil, the endemic form of PF (EPF) is popularly known as "*fogo selvagem*" (which means "wild fire") and reaches a highest prevalence of approximately 3.4% in an indigenous community living in the in Limão-Verde reservation area in the central-western region (Bystryn & Rudolph, 2005; Diaz et al., 1989 [2]; Hans-Filho et al., 1999). Moreover, both the endemic and sporadic forms of PF are clinically and immunologically similar, despite the increased IgM levels in EPF and endemic controls, suggesting a sensitization by environmental antigens in endemic areas (Culton et al., 2008; Diaz et al., 2008).

Regarding to the genetic aspects, certain alleles of human leukocyte antigen (HLA) class II genes (especially *HLA-DRB1* and *HLA-DQB1*), as well as other genes in the major histocompatibility complex (MHC), are highly associated with PV and PF (Petzl-Erler, 2020). Besides them, other genes are also associated with both diseases (Feliciani et al., 2000; Pereira et al., 2004; Eberhard et al., 2005; Pavoni et al., 2006; Braun-Prado & Petzl-Erler, 2007;

Malheiros & Petzl-Erler, 2009; Dalla-Costa et al., 2010; Mosaad et al., 2012; Augusto et al., 2012; Sarig et al., 2012; Piovezan & Petzl-Erler, 2013; Camargo et al., 2016; Tanasilovic et al., 2017; Salviano-Silva et al., 2017; Farias et al., 2018; Bumiller-Bini et al., 2018; Bumiller-Bini et al., 2019; Oliveira, L.C. et al., 2019; Spadoni et al., 2019).

Apart of protein-coding genes, the genetic influence of long non-coding RNAs (lncRNAs) have also been studied in PF (Lobo-Alves et al., 2019). The lncRNAs are large transcripts (more than 200 nucleotides) without protein-coding potential, involved in distinct mechanisms of transcriptional and post-transcriptional regulation. Therefore, the lncRNAs have been largely studied in many physiological and pathological processes (Salviano-Silva & Lobo-Alves et al., 2018; Cipolla et al., 2018; Oliveira, J.C. et al., 2019). Moreover, many lncRNAs are located in the MHC region, of which some lncRNAs of a multigene family classified as *HLA-Complex Group* (HCG) have also been associated with autoimmune and immune-related processes. Due to the importance of MHC to immunological and autoimmune responses (Kulsi, Shiina & Dijkstra, 2019; Tron et al., 2005), as well as the regulatory nature of lncRNAs, it is possible that HCG lncRNAs present an important role in pemphigus autoimmunity. However, there are no studies investigating the association of all HCG lncRNA genes in diseases, nor even of any HCG in pemphigus autoimmunity.

Aiming to explore the involvement of HCG lncRNAs in pemphigus susceptibility, we investigated if single nucleotide polymorphisms (SNPs) located in 13 HCG lncRNA genes are associated with PV and PF. Moreover, we evaluated the expression of the most associated HCG genes in patients PBMCs and cell culture. We found that HCG lncRNA genes are associated with both diseases, of which TSBP1-AS1 (which includes HCG23) and HCG27 are differentially expressed in PV and PF, respectively.

## 2. MATERIAL AND METHODS

### *2.1 Human samples*

This study was approved both by the National Committee for Ethics in Research (CONEP protocol CAAE 02727412.4.0000.0096, approval 505.988) and the Ethics Committee of the University of Lübeck (08-156, 12-178), according to Brazilian and German federal laws. All subjects voluntarily agreed to participate in this study and signed an informed consent form, following the Declaration of Helsinki. The study was performed in cooperation with the

Biobank PopGen, Kiel, Germany. All subjects were unrelated and with no clinical history of other autoimmune disease.

The PV cohort used for the genotyping study comprised 241 PV patients and 1188 controls, all of European origin. The subjects were collected in German hospitals by the German Autoimmune Bullous Diseases Genetic Study Group. PV patients were diagnosed based on a compatible clinical phenotype and a positive direct immunofluorescence microscopy of a perilesional skin biopsy or serum autoantibodies against Dsg3 as detected by ELISA (Euroimmun, Lübeck, Germany), according to the guideline of the German Society of Dermatology (Schmidt et al. 2010; Schmidt et al., 2015).

The endemic PF (EPF) cohort was composed by 227 PF patients and 194 controls from endemic areas in Brazil. All individuals, including the subjects enrolled in expression analysis (controls, EPF and PV), were of predominantly European ancestry, and were contacted at reference hospitals located in the endemic areas, as previously described (Cipolla et al., 2016). The patients were diagnosed based on physical examination combined with immunological testing, immunohistochemistry of skin biopsies and/or histopathology.

For all subjects, peripheral blood samples were collected and used for DNA isolations. Peripheral blood mononuclear cells (PBMCs) of Brazilian controls (7) and patients with PV (15, of which 8 with active disease) and EPF (6 with active disease) were also used for RNA quantifications. Moreover, serum of 5 PV patients, 5 sporadic PF patients and 6 control subjects from Germany were collected, for posterior purification of IgG antibodies used in cell culture.

Clinical and demographic data of subjects enrolled in this study are summarized in the Supplementary Table 1.

### *2.2 DNA samples and Microarray genotyping*

DNA was isolated from total blood samples of the PV cohort, using the Smart Blood DNA midi prep (Analytik Jena, Germany), and of the EPF cohort with phenol-chloroform-isoamyl alcohol method (Sambrook & Russell, 2001), and then stored at -80°C until genotyping.

For PV and EPF cohorts, the genotyping assays were performed with Global Screening Array and with Human Infinium® CoreExome-24 Beadchip (both of Illumina®, San Diego, USA), respectively, according to the manufacturer's instructions.

### *2.3 Selection of lncRNA candidates and data quality control*

We selected the 13 HCG lncRNA genes (**Figure 1**; **Table 1**) available in LNCiPedia DB (GRCh37/hg19) (Volders et al., 2019), for genetic variability investigation in pemphigus. For the *HCG23* gene, we selected the whole gene region of *TSBP1-AS1* ("TSBP1 and BTNL2 antisense RNA 1"), which includes HCG23 (*ENST00000426643.1*) (RNA Central DB - The RNAcentral Consortium, 2019; available in https://rnacentral.org/rna/URS000075B48F/9606) and whose name is still classified as HCG23 in some databases (LNCipedia DB, available in https://hg38.lncipedia.org/db/gene/HCG23). None of the selected genes were previously investigated in any subtype of pemphigus.

We searched for the genomic regions of the HCG genes at LNCiPedia DB (GRCh37/hg19) (Volders et al., 2019) in order to extract tag variants that mapped to these selected chromosome-6 genes, from the genotyping data of both PV and EPF microarrays. 554 genetic variants were extracted from PV microarray data, while 269 variants were extracted from data of EPF microarray. Manipulation of the SNP data was performed with PLINK version 1.09 (Chang et al., 2015). For the quality control, we excluded markers having (i) a minor allele frequency (MAF) < 5%, (ii) a genotype distribution deviating from Hardy-Weinberg equilibrium in the control sample ($p < 0.05$) and (iii) a strong linkage disequilibrium ($r^2 > 0.8$) with other SNPs in the analysis. A total of 160 SNPs remained for PV analysis, and 115 SNPs remained for EPF (Supplementary Table 2).

### *2.4 Computational analysis*

Computational analyses were performed to search functional annotations for all EPF-associated SNPs and for the 10 most associated SNPs with PV. Possible regulatory features and minor allele frequencies in European, African and Amerindian populations were searched in GRCh38/hg38 version of Ensembl genome browser (Howe et al., 2019). Due to the data divergences for some HCG genes in genomic databases, the SNP locations into their respective genes were searched in Ensembl genome browser (Howe et al., 2019), LNCipedia DB (Volders et al., 2019) and RNA Central DB (The RNAcentral Consortium, 2019). Possible eQTL

(expression quantitative loci trait) and sQTL (splice quantitative loci trait) effects in skin tissues were searched in GTEX Consortium project (Carithers et al., 2015). Molecular interactions data concerning the most associated genes with RNAs and proteins were searched in the ENCORI interactome database, as well as their participating pathways (LI et al., 2014).

All associated SNPs were evaluated for their linkage disequilibrium (LD) with other non-analyzed SNPs in their genomic extension, within population samples from the 1000 Genomes project data, using LDLink (Machiela & Chanock, 2017) online tool (available in https://ldlink.nci.nih.gov/?tab=ldassoc#). The searched populations were: CEU (Utah Residents with Northern and Western European Ancestry) for PV; and CEU, TSI (Toscani in Italia), IBS (Iberian Population in Spain) and YRI (Yoruba in Ibadan, Nigeria) for EPF.

### *2.5 Cell culture*

Keratinocytes (HaCaT human immortalized cell line) were grown in Keratinocyte Growth Medium 2 (Sigma-Aldrich, USA) at 37ºC and 5% $CO_2$, and seeded in different culture flasks, prior the treatment. IgG antibodies were purified from serum of 5 PV patients, 5 sporadic PF patients and 6 control subjects, all from Germany (Supplementary Table 1), using the Pierce™ Protein G IgG Agarose (Thermo Fischer Scientific, USA). After reaching ~80% of confluency, the HaCaT cells were stimulated with 1 mg/mL of IgG antibodies and incubated for 12 hours, prior to cell lysis for RNA isolations.

### *2.6 Quantification of RNA levels*

Total RNA was isolated from PBMCs with TRI Reagent® Solution (Thermo Fisher Scientific, USA), and cDNA samples were produced with High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, USA). For the HaCaT cells, large RNAs were isolated separately from small RNAs fraction using mirVana miRNA Isolation Kit (Thermo Fisher Scientific, USA), followed by reverse transcription with First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, USA). In reverse transcription assays, random primers were used for TSBP1-AS1 and HCG27, while specific reverse primers were used for HCG23 and TSBP1 RNAs.

Quantitative real-time polymerase chain reaction (qPCR) was used to measure the expression levels of *TSBP1-AS1, TSBP1, HCG23 and HCG27* genes, which were normalized by the housekeeping gene *RPL13A* (*Ribosomal Protein L13a*). The qPCR reactions were performed in

triplicates using Maxima SYBR Green/ROX qPCR Master Mix (Thermo Fisher Scientific, USA) and specific primer pairs **(Table 2)** in 2μL of cDNA. The qPCR amplifications were performed at 62ºC in StepOne Real-Time PCR Systems (Applied Biosystems®, USA). All amplifications presented unique peaks in the melting curves (data not shown).

### *2.7 Statistical analysis*

For association analysis, we performed a logistic regression using the additive model, with odds ratio (OR) and confidence interval of 95% (95%), applying corrections for sex and two principal components (PCs), in order to control for possible population structure.

Taking into account the high discrepancy in the quantity of associated SNPs found for EPF and PV, we adopted a significance limit of 0.05 for false discovery rate (FDR) adjusted *p* values in EPF (FDR *p* values between 0.05 and 0.1 were considered suggestive of association), and of 0.01 for FDR adjusted *p* values in PV (FDR *p* values between 0.01 and 0.05 were considered suggestive).

For the gene expression analysis, the fold changes (FC) were calculated by the comparative quantification cycle (Cq) method $2^{-\Delta\Delta Cq}$ and normalized by log2 (Livak & Schmittgen, 2001). *Mann-Whitney* tests were performed using GraphPad Prism v.6 software. The limit of significance adopted was $p < 0.05$.

## 3. RESULTS

### *3.1 SNPs in HCG lncRNA genes are associated with PV*

We found 55 HCG SNPs significantly associated with PV (FDR $p < 0.01$) and 22 suggestive of association ($0.01 <$ FDR $p < 0.05$), in the additive model **(Table 3)**. The associated SNPs are located in the following genes: *TSBP1-AS1* (38 SNPs, of which 2 are also in *HCG23*), *HCG17* (6 SNPs, of which one also overlaps *HCG18*, and another the *HLA-L* gene), *HCG27* (5 SNPs), *HCG22* (3 SNPs), *HCG20* (2 SNPs) and *HCG26/HCP5* (1 SNP). Among the associated SNPs in *TSBP1-AS1*, 30 are also located in the protein-coding gene *TSBP1*.The strongest association for PV was with *rs1003879*A*, that may increase the disease susceptibility (FDR $p = 5.24^{-22}$, OR= 4.78). This SNP locates in intron 4 of *TSBP1-AS1* that corresponds to intron 18 of *TSBP1* (Ensembl genome browser, Howe et al., 2019).

The suggestively associated SNPs occurred in *TSBP1-AS1* (15 SNPs, of which 10 are also in *TSBP1* coding gene), *HCG27* (3 SNPs), *HCG17* (2 SNPs, of which one is also located in *HCG18*), *HCG20* (1 SNP) and *HCG21* (1 SNP) genes **(Table 3)**.

These results lead us to suggest that the genetic variability of HCG lncRNA genes influences the susceptibility to PV.

### *3.2 SNPs in HCG lncRNA genes are associated with endemic PF*

We found 9 HCG SNPs significantly associated with EPF (FDR $p < 0.05$) and 6 suggestive of association ($0.05 <$ FDR $p < 0.1$), in the additive model **(Table 4)**. The associated SNPs are located in genomic regions of *TSBP1-AS1* (7 associated SNPs, of which 4 also occur in the *TSBP1* gene), *HCG27* (2 SNPs, especifically in the transcript HCG27:6) and *HCG21* (1 SNP, which also pertain to the *MUCL3* and *SFTA2* protein-coding genes). Interestingly, 3 EPF-associated SNPs (*rs3129943*G*, *rs3129949*A* and *rs9268103*A*) are also associated with PV, but with an opposite effect **(Tables 3 and 4)**. Increased susceptibility to EPF was strongly associated with *TSBP1-AS1 rs16870005*A* (FDR $p$ = 0.0135, OR=10.56). This SNP occurs in intron 4 of the *TSBP1-AS1* and in exon 30 of the *TSBP1* gene, as a missense variant for TSBP1 (**Table 4;** Supplementary Table 3).

The suggestively associated SNPs occurred in *TSBP1-AS1* (4 SNPs, of which 3 are also within *TSBP1*), *HCG27* (1 SNP) and *HCG20* (1 SNP) genes **(Table 4)**.

These results lead us to suggest that the genetic variability of HCG lncRNA genes influences the susceptibility to PF.

### *3.3 Functional annotations suggest potential relevance of TSBP1-AS1 and HCG27 variability*

Among the top-10 PV associations, the most associated SNP (*rs1003879*) is located in a motif with binding sites for many transcription factors (TF), considered by some genomic databases as a promoter flanking region for neighbor genes (Ensembl genome browser, Howe et al., 2019). Moreover, *rs1003879* does not present strong linkage disequilibrium (defined as $r^2 > 0.8$) with other SNPs, within the European-derived population of Utah (USA). Thus, the *rs1003879*A* allele may be a causal variant (Supplementary Table 3; Supplementary figure 1).

All of the PV most associated SNPs are within the intron 4 of *TSBP1-AS1*, excepting *rs2273017*G* (*TSBP1-AS1* intron 6; also located in a TF binding motif). Among them,

*rs3115562*T* (the second most associated SNP) is the unique SNP not overlapped with *TSBP1* coding gene, being specific of *TSBP1-AS1*. Furthermore, most of them have eQTL and sQTL effects for HLA and other MHC genes in skin tissues, being *rs521828*T* located in an acceptor splice site motif (Supplementary Table 3). Most SNPs present LD with other variants specifically in the gene region, with few exceptions of SNPs in LD with *HLA-DRA* variants (Supplementary Table 3; Supplementary figure 1).

Similarly to PV, the EPF associated SNPs in *TSBP1-AS1* are located in the intron 4 of this gene, excepting for a SNP in intron 6 (*rs3129943*G*, which is not in strong LD with the PV-associated *rs2273017*G* allele). Among them, the most associated SNP (*rs16870005*) is a missense variant responsible for the amino acid change from alanine to threonine, in the codon 431 of the canonical TSBP1 mRNA (Supplementary Table 3). Moreover, the EPF-associated SNPs are not strongly linked to HLA genes. While the *rs16870005*A* is in LD ($r^2 > 0.8$) with variants also located in neighbor genes, the other 8 associated SNPs only present LD with variants of their own genes, in all investigated populations (Supplementary Table 3; Supplementary figure 2). Most of these SNPs also show eQTL and sQTL effects in skin tissues (Supplementary Table 3).

Furthermore, the most associated genes, *TSBP1-AS1* and *HCG27*, are predicted to interact with various relevant RNAs and proteins, of which many interactions were found in analysis of many interactome sequencies data (ENCORI interactome database). The lncRNA *TSBP1-AS1* is capable to bind RNA binding proteins (RBPs) and also RNAs of different classes, including mRNAs and miRNAs. Among the miRNAs, the *TSBP1-AS1* can interact with the hsa-miR-552-3p, hsa-miR-3064-5p and hsa-miR-6504-5p, also acting as a competing endogenous RNA (ceRNA) for their targets and participating in many pathways (including pathways related to adhesion and immunological signals). Regarding to the *HCG27*, this lncRNA participates in the natural killer cell mediated cytotoxicity and p38 signaling pathways, as well as can interact with different transcripts of MIC genes and other mRNAs and RBPs molecules (ENCORI interactome database).

### *3.4 TSBP1-AS1 expression is low in PV's PBMCs*

We performed qPCRs for the genes with the most associated SNPs (*TSBP1-AS1*, *TSBP1*, *HCG23* and *HCG27*), to compare gene expression between PBMCs of pemphigus patients with and

without remission, and healthy individuals. We found a lower expression of *TSBP1-AS1* in PV patients with active disease, compared with control subjects (*p* = 0.029, log2 FC = -1.29). A trend for lower TSBP1-AS1 levels also occurred in patients under remission (*p* = 0.073, log2 FC = -1) **(Figure 2a)**. *HCG27* expression did not differ between controls and PV patients with active disease, but lower levels of this lncRNA occurred in patients under remission, in comparison with controls (*p* = 0.048, log2 FC = -2.19) **(Figure 2b)**. No significant or suggestive differences were found for HCG23 and TSBP1 RNA levels (data not shown).

These results lead us to suggest that the expression of *TSBP1-AS1* lncRNA is downregulated in PBMCs of PV patients, regardless of the current disease state, whereas HCG27 levels decreased in PV patients under remission, to a level below the one found in controls.

### *3.5 HCG27 is overexpressed in EPF PBMCs*

In contrast to PV, the expression of *HCG27* was obviously higher in PBMCs from EPF patients (*p* = 0.035, log2 FC = 1.3), while no significant differences were observed for *TSBP1-AS1* (**Figure 3a,b**). Instead, we found lower levels of TSBP1 mRNA (*p* = 0.042, log2 FC = -2.14), as well as suggestive slightly lower levels of HCG23 (*p* = 0.072, log2 FC = -1.99) in PBMCs of EPF patients (**Figure 3c,d**).

These results lead us to suggest that HCG27 lncRNA and TSBP1 mRNA are differentially expressed in EPF PBMCs.

### *3.6 TSBP1-AS1 and HCG27 are also differentially expressed in PV and PF IgG-stimulated keratinocytes*

As PV and PF are autoimmune diseases which affect the skin, we also aimed to investigate the expression of *TSBP1-AS1*, *TSPB1*, *HCG23* and *HCG27* in keratinocytes. To this end, we stimulated HaCaT cells with IgG antibodies isolated from the serum of patients (PV and sporadic PF) and controls, and evaluated gene expression by qPCR.

As in PBMCs, TSBP1-AS1 was downregulated in PV-stimulated HaCaT cells (*p* = 0.032, log2 FC = -0.99), while HCG27 was upregulated in PF-stimulated HaCaT cells (*p* = 0.032, log2 FC = 1.71), when compared to control cells (**Figure 4a,b**). In contrast to PBMCs, however, no significant differences were observed for TSBP1 and HCG23 in these keratinocyte samples (**Figure 4c,d**).

These results reinforce the occurrence of TSBP1-AS1 lower levels in PV, as well as HCG27 increased levels in both the endemic and non-endemic forms of PF.

## 4. DISCUSSION

The MHC is a high density gene region associated with many immunological processes and autoimmune responses (Kulsi, Shiina & Dijkstra, 2019), including pemphigus diseases, where some HLA alleles are the most associated genes (reviewed in Petzl-Erler, 2020). Still in the MHC region, the multigenic family of HCGs includes lncRNA genes, some of which have been associated with autoimmune and inflammatory disorders, especially in GWAS studies (GWAS catalog, Buniello et al., 2019), although rarely explored regarding to their expression and functions. Here, we aimed to investigate the genetic influence of HCG lncRNA members in pemphigus diseases.

First, we asked if the genetic variability of HCG lncRNAs influences the susceptibility for PV and EPF. For the best of our knowledge, this is the first work investigating the whole family of HCG lncRNA genes in a genetic disease association study. We found 55 HCG SNPs associated with PV and 9 associated with EPF, of which many are in the *TSBP1-AS1* gene, most also mapping to the *TSBP1* coding gene. Additional associations include SNPs located in: *HCG27:6*, also for both diseases; *HCG21* for EPF; and *HCG17*, *HCG18*, *HCG22*, *HCG20* and *HCG26* (overlapped with *HCP5* pseudogene) for PV. In contrast to the EPF associations, the increased amount of SNPs associated with PV (even adopting a more rigourous significance limit) might be consequence of a higher statistical power in this analyzed group, due to its higher controls sample size.

The highly associated gene *TSBP1-AS1* has approximately 139 kb, comprising *HCG23* (currently known as *TSBP1-AS1-204*, or *ENST00000426643.1*) and *HCG23:6* (*LOC101929163*, also considered a TSBP1-AS1 transcript). The TSBP1-AS1 functions are not yet elucidated, but it is known that this lncRNA is highly expressed in immune cells (Expression Atlas, Papatheodorou et al., 2020), and is genetically associated with many immune-related and dermatological diseases in GWAS studies (GWAS catalog, Buniello et al., 2019). *TSBP1-AS1* transcripts can also interact with many relevant molecules, such as RBPs, mRNAs and miRNAs (ENCORI interactome database). Indeed, *TSBP1-AS1* is pointed in interactome analysis as a ceRNA, by sponging the hsa-miR-552-3p, hsa-miR-3064-5p and hsa-miR-6504-5p miRNAs, thus possibly

influencing pathways related to cell adhesion and immunological signals, among others (ENCORI interactome database). Moreover, *TSBP1-AS1* partially overlaps the *TSBP1* (also known as *C6orf10*) and *BTNL2* protein-coding genes, which are transcribed from the opposite strand (LNCipedia DB; RNA Central DB), also being identified as pleiotropic genes associated with many autoimmune traits (Zheng et al., 2015; Anaparti et al., 2019; Gavrilova et al., 2019; Jin et al., 2011). In the present study, most *TSBP1-AS1* associated SNPs also pertain to the *TSBP1* gene, and 2 PV-associated SNPs are also located in the *HCG23* (*TSBP1-AS1*/*HCG23* SNPs: *rs17208671*T* and *rs3117098*G*/ *rs938761994*indel*).

The allele most associated with the increased susceptibility to EPF was *rs16870005*A*. Although intronic in the *TSBP1-AS1* gene, *rs16870005*A* is a missense variant for the *TSBP1*, resulting in an amino acid change from alanine to threonine in the codon 431 in the canonical *TSBP1* mRNA (Ensembl genome browser, Howe et al., 2019). Moreover, *rs16870005*A* has been associated with increased susceptibility to multiple sclerosis (Ziliotto et al., 2019). Regarding PV, *rs1003879*A,* an intronic *TSBP1-AS1/TSBP1* polymorphism, was strongly associated with the disease susceptibility. Due to the lack of strong LD with other SNPs, *rs1003879*A* is suggested as a possible causal SNP in PV pathogenesis.

For both diseases, the most associated SNPs are mainly located within the intron 4 of *TSBP1-AS1*, a region which overlaps the end of the *TSBP1* gene and might host and/or be in LD with important regulatory regions. Indeed, the PV-associated *rs1003879* and the EPF-associated *rs3132931* are in motifs of TF binding sites, while the PV-associated *rs521828* is a splice acceptor variant for *TSBP1-AS1* (Ensembl genome browser, Howe et al., 2019). The PV-associated SNPs *rs2273017* and *rs3129943* are also located in TF binding sites, but in intron 6 of *TSBP1-AS1* (first exons of *TSBP1*). Interestingly, the *rs3129943*G*, as well as *rs3129949*A* and *rs9268103*A* are associated with both PV and EPF diseases, however, with opposite effects. In addition, most of these SNPs have eQTL and sQTL effects in the skin, for their own genes and for other MHC genes, including the pemphigus-associated *HLA-DRB1* and *HLA-DQB1* (GTEX Portal DB). Among the PV protective associations, the *TSBP1-AS1 rs7775397*G* was associated with increased neonatal lupus transmission (Saxena et al., 2011), while the *rs2395149* and *rs1265757* were associated with increased susceptibility to Grave's disease (Khong et al., 2016).

Considering these results, we aimed to evaluate the expression profile of the associated genes in patients and controls. We found lower *TSBP1-AS1* levels in PBMCs of Brazilian PV patients, as well as in cultured keratinocytes stimulated with IgG antibodies of German PV patients, when compared with their respective control samples. The absence of differential expression for *TSBP1* leads us to suggest that *TSBP1-AS1* lncRNA plays a causal role for PV pathogenesis. Regarding to the interaction abilities between *TSBP1-AS1* and other relevant molecules (ENCORI interactome database), it is possible that the lower levels of *TSBP1-AS1* could influence on the expression and availability of those molecules, with consequences in signaling pathways involved with pemphigus autoimmunity.

On the other hand, while no expression differences were observed for *TSBP1-AS1* in PF, we found *TSBP1* mRNA levels decreased in EPF PBMCs, but not in keratinocytes treated with sporadic PF IgG antibodies. Furthermore, non-significant differences were observed for *HCG23* in PF samples, whose expression seems to be slightly decreased in EPF PBMCs and slightly increased in sporadic PF-stimulated keratinocytes. This discrepancy could be solely due to the physiological expression differences in both cell types or could even indicate that *HCG23* is differentially expressed between sporadic and endemic forms, with biomarker potential for differentiating both PF clinical forms. However, further studies concerning different subgroups with larger sample sizes must be performed to investigate this hypothesis, as well as possible explanations underlying the *TSBP1* and *HCG23* differential expressions among pemphigus groups.

As the second most associated gene in this study, *HCG27:6* has one associated SNP with EPF and 5 SNPs associated with PV. The *HCG27* is an intergenic lncRNA gene with approximately 31 kb, located in a genomic region associated with psoriasis, an inflammatory disease also affecting the skin (Clop et al., 2013; GWAS catalog, Buniello et al., 2019). Moreover, differential levels of *HCG27* were already found in type II diabetes mellitus, also being correlated to disease-associated SNPs (Saeidi et al., 2018). Here, we also found *HCG27* overexpressed in PBMCs of Brazilian EPF patients and in cultured keratinocytes stimulated with IgG antibodies of German sporadic PF patients, in relation to controls. These results suggest that increased levels of this lncRNA might be related with PF autoimmunity, for both endemic and sporadic forms. Moreover, although no *HCG27* differential levels were found in patients with active PV, this lncRNA was decreased in PV patients under remission.

Considering the possible *HCG27* interactions with important RBPs and immune-related mRNAs (such as *MICA* and *MICB*) (ENCORI interactome database), as well as with pathways previously associated with pemphigus (such as natural killer cell mediated cytotoxicity and p38 signaling pathways) (ENCORI interactome database; Augusto et al., 2012, 2015; Cipolla et al. 2017, Stern et al., 2008), we suggest that those *HCG27* differential levels might bind and influence the availability of relevant target molecules, influencing in the regulation of immune-related genes and leading to the autoimmune process. However, such functional hypothesis might be explored by additional investigations (concerning interactions and loss of function studies, among others) in pemphigus diseases.

Among the methodological limitations of this study, the lncRNA SNPs in the analyzed genomic region were not fully covered by the genotyping microarrays. Moreover, due to microarray chip specifications and genetic population differences (regarding to SNP frequencies and LD data, which in turn were used for the quality controls), most remaining SNPs for statistical analysis were not the same for PV and EPF cohorts. Besides, we did not compare the HCG levels in PBMC samples according to the subject's genotyping, focusing our expression analysis in a case control approach. Nevertheless, we found genes associated with both diseases, whose expressions were demonstrated to be differential in patients and controls, both in PBMCs and in functional assays with keratinocytes, leading us to suggest for the first time, a potential role of HCG lncRNA genes in pemphigus diseases.

It is suggested that around 90% of causal autoimmune disease variants are located in non protein-coding regions, of which a significant part maps to immune-cell enhancer elements, many transcribing for enhancer non-coding RNAs (Farh et al., 2015). Therefore, although the still unknown mechanisms underlying the functional relevance of HCG associated SNPs in pemphigus diseases, it is possible that they are located in regulatory regions related to immune modulation, thus acting as enhancers for immune stimulation. Indeed, apart their interaction abilities with other molecules, evidencies also point for the relevance of the of lncRNA genomic locations, as they can act *in cis* to regulate the expression of neighbor genes (Toiber et al., 2017). Consequently, the characterization of functional non-coding genes within the MHC region, such as the HCG lncRNAs, might contribute to elucidate the mechanisms underlying immune-regulation of MHC genes in health and disease. Therefore, the herein results and hyphotesis should be further investigated in different cohorts and detailed

explored regarding to their molecular functions and regulatory aspects, as well as in the lymphocytic context.

Taking together, we conclude that genetic variants of HCG lncRNAs are associated with both PV and EPF, being *TSBP1-AS1* and *HCG27* aberrantly expressed in the respective diseases, in different population samples.  In addition, this study reinforces the importance of the MHC region in pemphigus autoimmunity and presents novel insights of lncRNAs in this disease group, suggesting HCG lncRNAs as powerful candidate genes for exploratory studies of genetic effects and biomarkers in PV and PF.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the patients for their participation in this study. We also thank the staff of the Laboratório de Genética Molecular Humana/UFPR for their assistance and helpful discussions, as well as the investigators of the German Autoimmune Bullous Diseases Study Group.

## FUNDING SOURCES

This work was supported by grants of Fundação Araucária (FA), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Finance Code 001, , Co-financed Short-Term Research Grant Brazil (2018) by Deutscher Akademischer Austauschdienst (DAAD), and the German Research Foundation under Germany’s Excellence Strategy (EXC 2167-390884018; HB, DZ, ES).

## CONFLICT OF INTEREST

The authors declare no conflicts of interest associated with this publication.

## REFERENCES

Anaparti, V., Agarwal, P., Smolik, I., Mookherjee, N., & Elgabalawy, H. (2019). Whole Blood Targeted Bisulfite Sequencing Validates Differential Methylation in C6ORF10 gene of Patients with Rheumatoid Arthritis. **The Journal of Rheumatology**, jrheum.190376. doi:10.3899/jrheum.190376

Augusto, D. G.; Lobo-Alves, S. C.; Melo, M. F.; Pereira, N. F.; Petzl-Erler, M. L. Activating KIR and HLA Bw4 ligands are associated to decreased susceptibility to pemphigus foliaceus, an autoimmune blistering skin disease. **PLoS One**, v. 7 (7), 2012.

Augusto DG, O'Connor GM, Lobo-Alves SC, et al. Pemphigus is associated with KIR3DL2 expression levels and provides evidence that KIR3DL2 may bind HLA-A3 and A11 in vivo. **Eur J Immunol.** 2015;45(7):2052–2060. doi:10.1002/eji.201445324

Braun-Prado, K.; Petzl-Erler, M. L. Programmed cell death 1 gene (PDCD1) polymorphism and pemphigus foliaceus (fogo selvagem) disease susceptibility. **Genetics and Molecular Biology**, v. 30, p. 314–321, 2007.

Brochado MJ, Nascimento DF, Campos W, Deghaide NH, Donadi EA, Roselino AM. Differential HLA class I and class II associations in pemphigus foliaceus and pemphigus vulgaris patients from a prevalent Southeastern Brazilian region. J Autoimmun. 2016 Aug;72:19-24. doi: 10.1016/j.jaut.2016.04.007. Epub 2016 May 10.

Bumiller-Bini, V.; Cipolla, G.A.; De Almeida, R.C.; Petzl-Erler, M.L.; Augusto, D.G.; Boldt, A.B.W. Sparking Fire Under the Skin? Answers From the Association of Complement Genes With Pemphigus Foliaceus. **Front Immunol.** V.9:695, 2018.

Bumiller-Bini, V.; Cipolla, G.A.; Spadoni, M.B.; Augusto, D.G.; Petzl-Erler, M.L.; Beltrame, M.H.; Boldt, A.B.W. Condemned or Not to Die? Gene Polymorphisms Associated With Cell Death in Pemphigus Foliaceus. **Front. Immunol.,** 18 October 2019 | https://doi.org/10.3389/fimmu.2019.02416

Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousgou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorff LA, Cunningham F and Parkinson H. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. **Nucleic Acids Research**, Vol. 47 (Database issue): D1005-D1012, 2019.

Bystryn, J. C.; Rudolph, J. L. Pemphigus.**The Lancet**, v. 366 (9479), p. 61–73, 2005.

Camargo, C. M.; Augusto, D. G.; Petzl-Erler, M. L. Differential gene expression levels might explain association of LAIR2 polymorphisms with pemphigus. **Human Genetics**, v. 135 (2), p. 233-244, 2016.

Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. **Biopreserv Biobank** 2015; 13:311–9.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. **Gigascience** 2015; 4:7.

Cipolla, G. A. et al. A 3’UTR polymorphism marks differential KLRG1 mRNA levels through disruption of a miR-584-5p binding site and associates with pemphigus foliaceus susceptibility. **Biochimica et Biophysica Acta**, v. 1859, p. 1306-1313, 2016.

Cipolla GA, Park JK, Lavker RM, Petzl-Erler ML. Crosstalk between Signaling Pathways in Pemphigus: A Role for Endoplasmic Reticulum Stress in p38 Mitogen-Activated Protein Kinase Activation?. **Front Immunol.** 2017;8:1022. Published 2017 Sep 5. doi:10.3389/fimmu.2017.01022

Cipolla, G. A., DE OLIVEIRA, J. C., SALVIANO-SILVA, A., LOBO-ALVES, S. C., LEMOS, D. S., OLIVEIRA, L. C., JUCOSKI, T. S., MATHIAS, C., PEDROSO, G. A., ZAMBALDE, E. P., GRADIA, D. F. Long Non-Coding RNAs in Multifactorial Diseases: Another Layer of Complexity. **Non-coding RNA**, 4(2), 13, 2018.

Clop, A., Bertoni, A., Spain, S. L., Simpson, M. A., Pullabhatla, V., Tonda, R., … Trembath, R. C. (2013). An in-depth characterization of the major psoriasis susceptibility locus identifies candidate susceptibility alleles within an HLA-C enhancer element. **PloS one**, 8(8), e71690. doi:10.1371/journal.pone.0071690

Culton, D. A.; Qian, Y.; Li, N.; Rubenstein, D.; Aoki, V.; Filho, G. H.; Rivitti, E. A.; Diaz, L. A. Advances in Pemphigus and its Endemic Pemphigus Foliaceus (FogoSelvagem) Phenotype: A Paradigm of Human Autoimmunity. **Journal of Autoimmunity**, v. 31 (4), p. 311–324, 2008.

Dalla-Costa R. et al. Polymorphisms in the 2q33 and 3q21 chromosome regions including Tcell coreceptor and ligand genes may influence susceptibility to pemphigus foliaceus. **Human Immunology**, v. 71, p. 809, 2010.

Diaz, L. A.; SAMPAIO, S. A.; RIVITTI, E. A.; MARTINS, C. R.; CUNHA, P. R.; LOMBARDI, C.; ALMEIDA, F. A.; CASTRO, R. M.; MACCA, M. L.; LAVRADO, C.; FILHO, G. H.;BORGES, P.; MINELLI, L.; EMPINOTTI, J. C.; FRIEDMAN, H.; CAMPBELL, I.; LABIB, R. S.;ANHALT, G. J. Endemic pemphigus foliaceus (fogo selvagem). I. Clinical features and immunopathology. **Journal of the American Academy of Dermatology**, V. 20 (4), P. 657–669, 1989.

Diaz, L. A.; SAMPAIO, S. A.; RIVITTI, E. A.; MARTINS, C. R.; CUNHA, P. R.; LOMBARDI, C.; ALMEIDA, F. A.; CASTRO, R. M.; MACCA, M. L.; LAVRADO, C.; FILHO, G. H.;BORGES, P.; CHAUL, A.; MINELLI, L.; EMPINOTTI, J. C.; FRIEDMAN, H.; CAMPBELL, I.; LABIB, R. S.;ANHALT, G. J. Endemic Pemphigus Foliaceus (Fogo Selvagem): II. Current and Historic Epidemiologic Studies. **Journal of Investigative Dermatology**, v. 92, p. 4–12, 1989. [2]

Diaz LA1, Prisayanh PS, Dasher DA, Li N, Evangelista F, Aoki V, Hans-Filho G, dos Santos V, Qaqish BF, Rivitti EA; Cooperative Group on Fogo Selvagem Research. The IgM anti-desmoglein 1 response distinguishes Brazilian pemphigus foliaceus (fogo selvagem) from other forms of pemphigus. **J Invest Dermatol**. 2008 Mar;128(3):667-75. Epub 2007 Oct 25.

Eberhard Y1, Burgos E, Gagliardi J, Vullo CM, Borosky A, Pesoa S, Serra HM. Cytokine polymorphisms in patients with pemphigus. **Arch Dermatol Res**., Jan;296(7):309-13, 2005. Epub 2004 Dec 10.

Farh, K., Marson, A., Zhu, J. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. **Nature** 518, 337–343 (2015). https://doi.org/10.1038/nature13835

Farias, T.D.J.; Augusto, D.G.; De Almeida, R.C.; Malheiros, D.; Petzl-Erler, M.L. Screening the full leucocyte receptor complex genomic region revealed associations with pemphigus that might be explained by gene regulation. **Immunology**, 156(1):86-93, 2019.

Feliciani C, Toto P, Amerio P, Pour SM, Coscione G, Shivji G, et al. . In vitro and in vivo expression of interleukin-1alpha and tumor necrosis factor-alpha mRNA in pemphigus vulgaris: interleukin-1alpha and tumor necrosis factor-alpha are involved in acantholysis. **J Invest Dermatol**., 114:71–7, 2000. 10.1046/j.1523-1747.2000.00835.x

Gavrilova T. Primary Immunodeficiency with Severe Multi-Organ Immune Dysregulation. **Case Reports Immunol.** 2019;2019:8746249. Published 2019 Dec 28. doi:10.1155/2019/8746249

Hans-Filho, G.; Aoki, V.; Rivitti, E. A.; Eaton, D. P.; Mong, S. L.; Diaz, L. A. The cooperative group on fogo selvagem research endemic pemphigus foliaceous (fogo selvagem). **Clinical Dermatology**, v. 17, p. 225-235, 1999.

Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Allen J, Alvarez-Jarreta J, Barba M, Bolser DM, Cambell L, Carbajo M, Chakiachvili M, Christensen M, Cummins C, Cuzick A, Davis P, Fexova S, Gall A, George N, Gil L, Gupta P, Hammond-Kosack KE, Haskell E, Hunt SE, Jaiswal P, Janacek SH, Kersey PJ, Langridge N, Maheswari U, Maurel T, McDowall MD, Moore B, Muffato M, Naamati G, Naithani S, Olson A, Papatheodorou I, Patricio M, Paulini M, Pedro H, Perry E, Preece J, Rosello M, Russell M, Sitnik V, Staines DM, Stein J, Tello-Ruiz MK, Trevanion SJ, Urban M, Wei S, Ware D, Williams G, Yates AD, Flicek P. Ensembl Genomes 2020-enabling non-vertebrate genomic research. **Nucleic Acids Research** 2019. https://doi.org/10.1093/nar/gkz890

Hrdlickova B, Kumar V, Kanduri K, et al. Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity. **Genome Med.** 2014;6(10):88. Published 2014 Oct 28. doi:10.1186/s13073-014-0088-0.

Irene Papatheodorou, Pablo Moreno, Jonathan Manning, Alfonso Muñoz-Pomer Fuentes, Nancy George, Silvie Fexova, Nuno A Fonseca, Anja Füllgrabe, Matthew Green, Ni Huang, Laura Huerta, Haider Iqbal, Monica Jianu, Suhaib Mohammed, Lingyun Zhao, Andrew F Jarnuczak, Simon Jupp, John Marioni, Kerstin Meyer, Robert Petryszak, Cesar Augusto Prada Medina, Carlos Talavera-López, Sarah Teichmann, Juan Antonio Vizcaino, Alvis Brazma. Expression Atlas update: from tissues to single cells. **Nucleic Acids Research**, V. 48, Issue D1, P. D77–D83, 2020. https://doi.org/10.1093/nar/gkz947.

Jin Y, Birlea SA, Fain PR, et al. Genome-wide analysis identifies a quantitative trait locus in the MHC class II region associated with generalized vitiligo age of onset. **J Invest Dermatol.** 2011;131(6):1308–1312. doi:10.1038/jid.2011.12

Kasperkiewicz, M., et al., Pemphigus. **Nat Rev Dis Primers**, 2017. 3: p. 17026.

Khong, J. J., Burdon, K. P., Lu, Y., Laurie, K., Leonardos, L., Baird, P. N., … Craig, J. E. (2016). Pooled genome wide association detects association upstream of FCRL3 with Graves' disease. **BMC genomics**, 17(1), 939. doi:10.1186/s12864-016-3276-z.

Kulski, J.K.; Shiina, T.; Dijkstra, J.M. Genomic Diversity of the Major Histocompatibility Complex in Health and Disease. Cells, 8(10) 1270, 2019. https://doi.org/10.3390/cells8101270

Li JH, et al.starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data , **Nucleic Acids Res**. 2014 Jan;42:D92-7.

Li, S.; Zhang, Q.; Wang, P.; Li, J.; Ni, J.; Wu, J.; Liang, Y.; Leng, R.X.; Pan, H.F.; Ye, D.Q. Association between HLA-DQB1 polymorphisms and pemphigus vulgaris: A meta-analysis. **Immunol Invest**., Jan;47(1):101-112, 2018. doi: 10.1080/08820139.2017.1385622. Epub 2017 Nov 28.

Livak, K.; Schmittgen, T. D. Analysis of relative gene expression data using realtime quantitative PCR and the 2-ΔΔCt method. **Methods**, v.25, p. 402-408, 2001.

Lobo-Alves, S.C.; Augusto, D.G.; Magalhães, W.C.S.; Tarazona-Santos, E.; Lima-Costa, M.F.; Barreto, M.L.; Horta, B.L.; De Almeida, R.C.; Petzl-Erler, M.L. lncRNA polymorphisms in EPF. **British Journal of Dermatology**, 181, 324–331, 2019. https://doi.org/10.1111/bjd.17640

Lombardi, M.L.; Mercuro, O.; Ruocco, V.; Lo Schiavo, A.; Lombari, V.; Guerrera, V.; Pirozzi, G.; Manzo, C. Common human leukocyte antigen alleles in pemphigus vulgaris and pemphigus foliaceus Italian patients. **J Invest Dermatol**. Jul;113(1):107-10, 1999.

Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. **Bioinformatics** 2015; 31:3555–7.

Malheiros, F. D.; Petzl-Erler, M. L. Individual and epistatic effects of genetic polymorphisms of B-cell co-stimulatory molecules on susceptibility to pemphigus foliaceus. **Genes and Immunity**, v. 10, p. 547–558, 2009.

Mosaad YM, Fathy H, Fawzy Z, El-Saied MA. Tumor necrosis factor-alpha−308 G>A and interleukin-6−174 G>C promoter polymorphisms and pemphigus. **Hum Immunol.** (2012) 73:560–5. 10.1016/j.humimm.2012.02.001

Nisihara, R. M.; De Bem, R. S.; Hausberger, R.; Roxo, V. S.; Pavoni, D. P.; Petzl-Erler, M. L.; Messias-Reason, I. J. Prevalence of autoantibodies in patients with endemic pemphigus foliaceus (fogo selvagem). **Archives of Dermatological research**, v. 295, p. 133-137, 2003.

Oliveira J.C.; Oliveira, L.C.; Mathias, C; Pedroso, G.A.; Lemos, D.S.; Salviano-Silva, A.; Jucoski, T.S.; Lobo-Alves, S.C.; Zambalde. E.P.; Cipolla, G.A.; Gradia, D.F. Long non-coding RNAs in cancer: Another layer of complexity. **The Journal of Gene Medicine**, V.21, 2019.

Oliveira, L. C.; Kretzschmar, G. C.; Dos Santos, A.; Camargo, C. M.; Nisihara, R. M.; Farias, T.; … ; Boldt, A. B. W. Complement Receptor 1 (CR1, CD35) Polymorphisms and Soluble CR1: A Proposed Anti-inflammatory Role to Quench the Fire of "Fogo Selvagem" Pemphigus Foliaceus. **Front. Immunol.**, 22 November 2019 | https://doi.org/10.3389/fimmu.2019.02585

Pavoni DP, Cerqueira LB, Roxo VM, Petzl-Erler ML. Polymorphism of the promoter region and exon 1 of the CTLA4 gene in endemic pemphigus foliaceus (fogo selvagem). **Braz J Med Biol Res**. 2006 Sep;39(9):1227-32.

Pavoni, D. P.; Roxo, V. M. M. S.; Marquart Filho, A.; Petzl-Erler, M. L. Dissecting the associations of endemic Pemphigus Foliaceus (FogoSelvagem) with HLA-DRB1 alleles and genotypes. **Genes and Immunity**, v. 4, p. 110–116, 2003.

Pereira, N. F.; Hansen, J. A.; Lin, M. T.; Roxo, V. M. M. S.; Braun, K.; Petzl-Erler, M. L. Cytokine gene polymorphisms in endemic pemphigus foliaceus: a possible role for IL6 variants. **Cytokine**, v. 28, p. 233-241, 2004.

Petzl-Erler, M. L.; Santamaria, J. Are HLA class II genes controlling susceptibility and resistance to Brazilian pemphigus foliaceus (fogo selvagem)? **Tissue Antigens**, v. 33, p. 408–414, 1989.

Petzl-Erler. Beyond the HLA polymorphism: A complex pattern of genetic susceptibility to pemphigus. **Genetics and Molecular Biology**, 2020 (accepted for publication).

Piovezan B.Z., Petzl-Erler M.L. Both Qualitative and quantitative genetic variation of MHC Class II molecules may influence susceptibility to autoimmune diseases: the case of endemic pemphigus foliaceus. **Human Immunology**, v.74, p.1134-1140, 2013.

Pollmann, R., et al., Pemphigus: a Comprehensive Review on Pathogenesis, Clinical Presentation and Novel Therapeutic Approaches. **Clin Rev Allergy Immunol**, 2018. 54(1): p. 1-25.

The RNAcentral Consortium. RNAcentral: a hub of information for non-coding RNA sequences [published correction appears in Nucleic Acids Res. 2019 Jan 8;47(D1):D1250-D1251]. Nucleic Acids Res. 2019;47(D1):D221–D229. doi:10.1093/nar/gky1034

Saeidi, L., Ghaedi, H., Sadatamini, M., Vahabpour, R., Rahimipour, A., Shanaki, M., … Kazerouni, F. (2018). Long non-coding RNA LY86-AS1 and HCG27_201 expression in type 2 diabetes mellitus. **Molecular Biology Reports**. doi:10.1007/s11033-018-4429-8

Salviano-Silva, A.; Petzl-Erler, M. L; Boldt, A.B.W. CD59 polymorphisms are associated with gene expression and different sexual susceptibility to pemphigus foliaceus. **Autoimmunity**, v50(6), 377-385, 2017.

Salviano-Silva, A., Lobo-Alves, S. C., Almeida, R. C., Malheiros, D., Petzl-Erler, M. L. Besides Pathology: Long Non-Coding RNA in Cell and Tissue Homeostasis. **Non-coding RNA**, 4(1), 3, 2018.

Sambrook J, Russell DW. Molecular Cloning: A laboratory manual. New York: CSHL Press, 2001

Sarig O, Bercovici S, Zoller L, Goldberg I, Indelman M, Nahum S, et al. . Population-specific association between a polymorphic variant in ST18, encoding a pro-apoptotic molecule, and pemphigus vulgaris. J Invest Dermatol., 132:1798–805, 2012. 10.1038/jid.2012.46

Saxena, A., McDonnell, E., Ramos, P. S., Sajuthi, S., Marion, M. C., Langefeld, C. D., … Clancy, R. M. (2012). Preferential transmission of genetic risk variants of candidate loci at 6p21 from asymptomatic grandparents to mothers of children with neonatal lupus. **Arthritis and rheumatism**, 64(3), 931–939. doi:10.1002/art.33366

Schmidt, E., M. Kasperkiewicz, and P. Joly, Pemphigus. **Lancet**, in press.

Schmidt E, Dahnrich C, Rosemann A et al. Novel ELISA systems for antibodies to desmoglein 1 and 3: correlation of disease activity with serum autoantibody levels in individual pemphigus patients. **Exp Dermatol** 2010; 19: 458-63.

Schmidt E, Goebeler M, Hertl M et al. S2k guideline for the diagnosis of pemphigus vulgaris/foliaceus and bullous pemphigoid. **J Dtsch Dermatol Ges** 2015; 13: 713-27.

Sinha, A. A. The Genetics of Pemphigus. **Dermatologic Clinics**, v. 29 (3), p. 381–391, 2011.

Spadoni, M.B.; Bumiller-Bini, V.; Petzl-Erler, M.L.; Augusto, D.G.; Boldt, A.B.W. First Glimpse of Epigenetic Effects on Pemphigus Foliaceus. **J Invest Dermatol.**, Feb;140(2):488-491.e1, 2020. doi: 10.1016/j.jid.2019.07.691. Epub 2019 Jul 31.

Spindler, V., et al., Mechanisms Causing Loss of Keratinocyte Cohesion in Pemphigus. **J Invest Dermatol**, 2018. 138(1): p. 32-37.

Stern JN, Keskin DB, Barteneva N, Zuniga J, Yunis EJ, Ahmed AR. Possible role of natural killer cells in pemphigus vulgaris - preliminary observations. **Clin Exp Immunol.** 2008;152(3):472–481. doi:10.1111/j.1365-2249.2008.03638.x

TAKEICHI, M. Cadherin cell adhesion receptors as a morphogenetic regulator. **Science**, v. 251 (5000), p. 1451-1455, 1991.

Tanasilovic S, Popadic S, Medenica L, Popadic D. Pemphigus vulgaris and pemphigus foliaceus determined by CD86 and CTLA4 polymorphisms. **Clin Dermatol.**, 35:236–41, 2017. 10.1016/j.clindermatol.2016.05.021

Toiber, D.; Leprivier, G.; & Rotblat, B. Long noncoding RNA: noncoding and not coded. **Cell Death Discovery**, 3, 16104, 2017. doi:10.1038/cddiscovery.2016.104

Tron, F.; Gilbert, D.; Mouquet, H.; Joly, P.; Drouot, L.; Makni, S.; Masmoudi, H.; Charron, D.; Zitouni, M.; Loiseau, P.; Ben Ayed, M. Genetic factors in pemphigus. **J Autoimmun**., Jun;24(4):319-28, 2005.

Vodo, D., Sarig, O., & Sprecher, E. The Genetics of Pemphigus Vulgaris. Frontiers in medicine, 5, 226, 2018. doi:10.3389/fmed.2018.00226

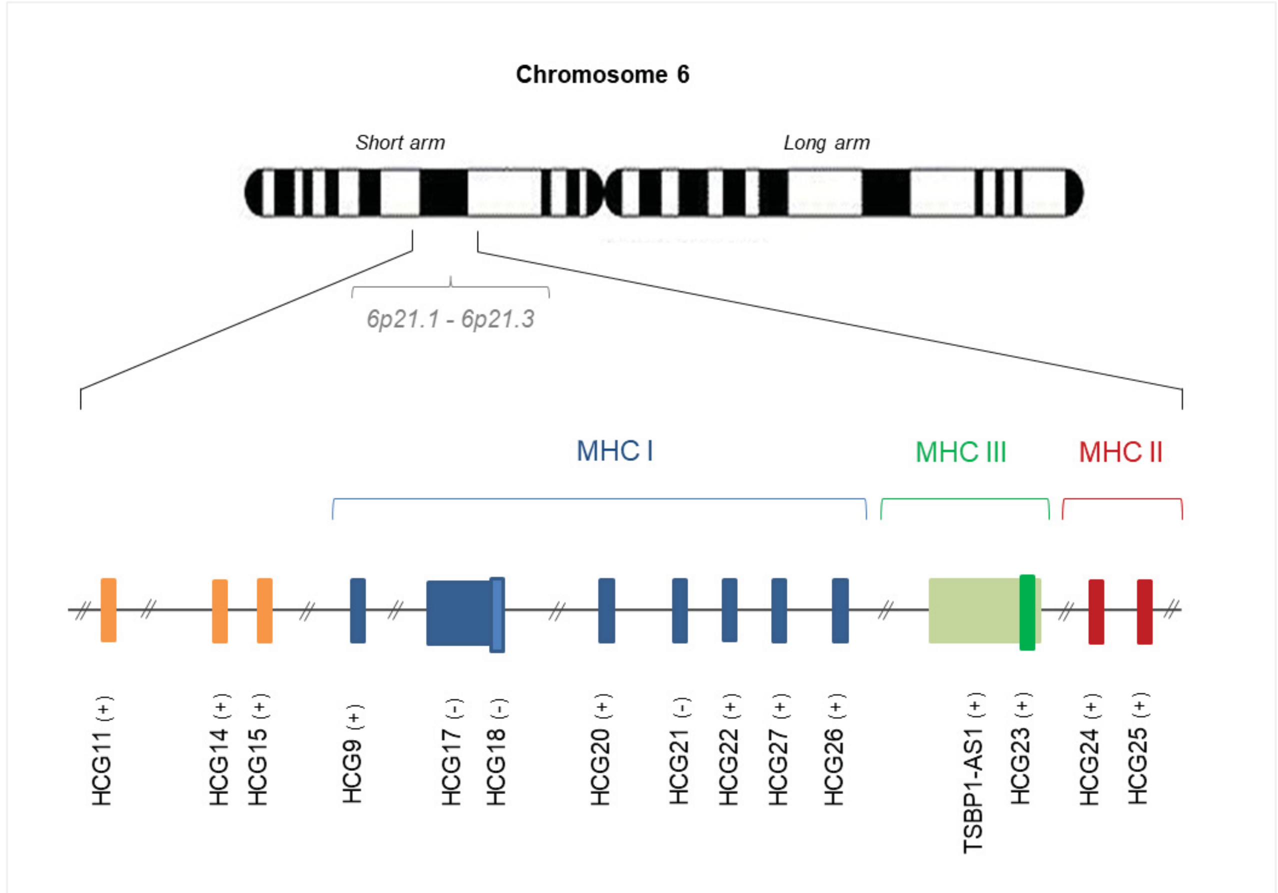
Pieter-Jan Volders; Jasper Anckaert; Kenneth Verheggen; Justine Nuytens; Lennart Martens; Pieter Mestdagh and Jo Vandesompele. LNCipedia 5: towards a reference set of human long non-coding RNAs. **Nucleic Acids Research** 2019. doi:10.1093/nar/gky1031.

Zheng W, Rao S. Knowledge-based analysis of genetic associations of rheumatoid arthritis to inform studies searching for pleiotropic genes: a literature review and network analysis. **Arthritis Res Ther**. 2015;17(1):202. Published 2015 Aug 8. doi:10.1186/s13075-015-0715-1

Ziliotto, Nicole et al. "C6orf10 Low-Frequency and Rare Variants in Italian Multiple Sclerosis Patients." **Frontiers in genetics** vol. 10 573. 26 Jun. 2019, doi:10.3389/fgene.2019.00573

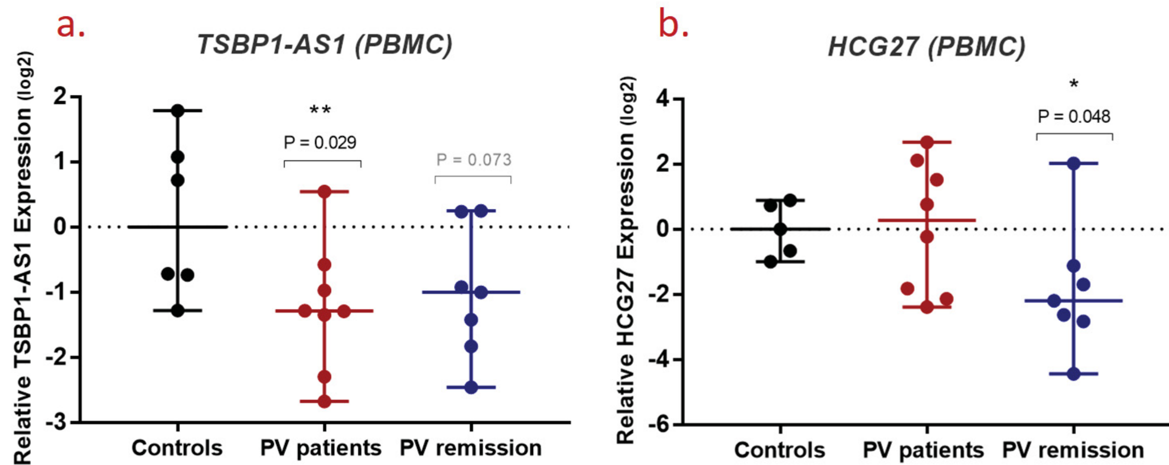
## *FIGURES AND TABLES*

**Figure 1.** Genomic representation of HCG lncRNA genes in the chromosome 6.



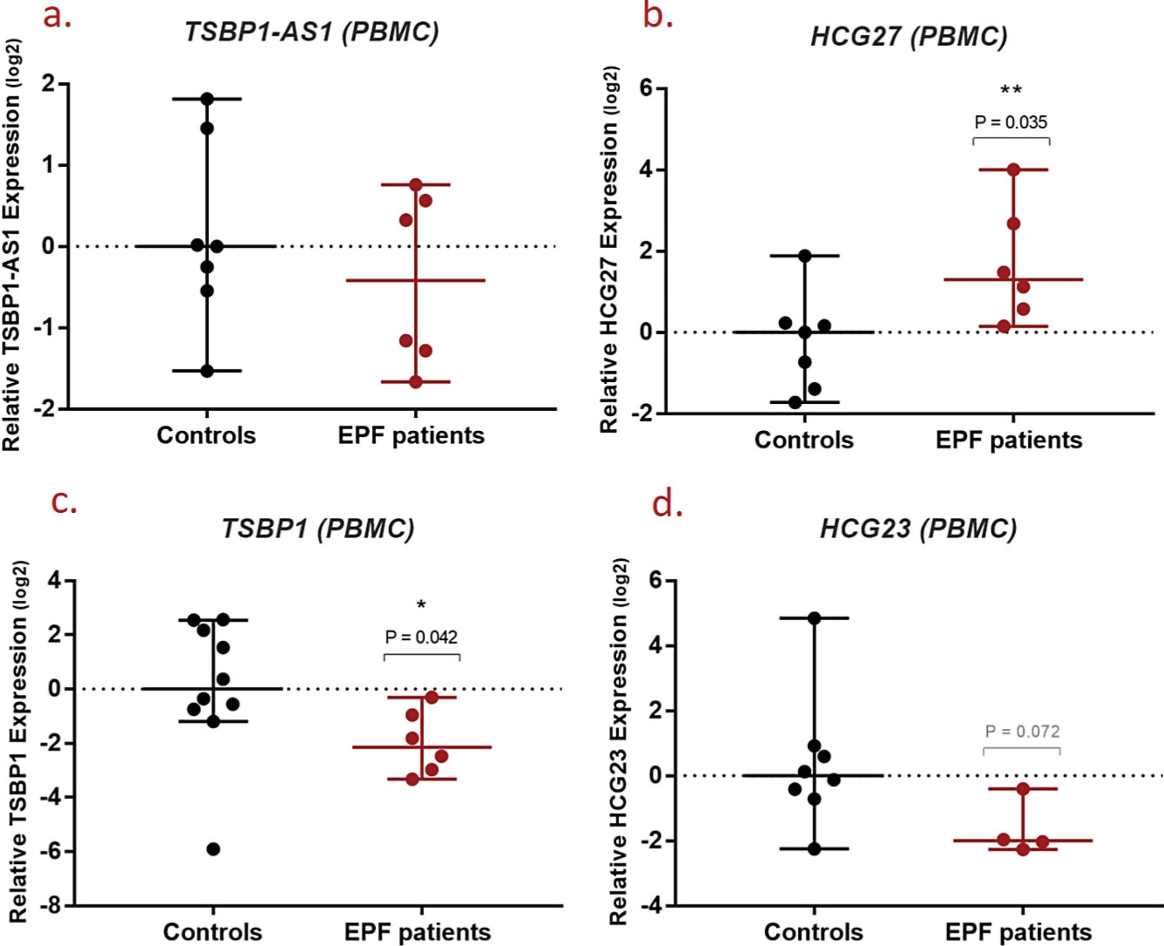

In Orange, there are represented the HCG lncRNA genes located in the region 6p.21.1-21.2, upstream to the MHC I (which starts in 6p21.3); in blue, there are represented the HCG genes located in MHC class I; in green, the HCG23 gene located in MHC class III, included in TSBP1-AS1 lncRNA; in red, the HCG lncRNA genes located in the MHC class II. Other overlapped genes are no represented in this figure. The genomic locations are merely illustrative, not representing the real scale of the region. The symbols "(+)" and "(-)" indicate the strand orientation where the gene is inserted (forward or reverse, respectively).

**Figure 2.** TSBP1-AS1 and HCG27 RNA levels in PBMCs of PV patients and controls.



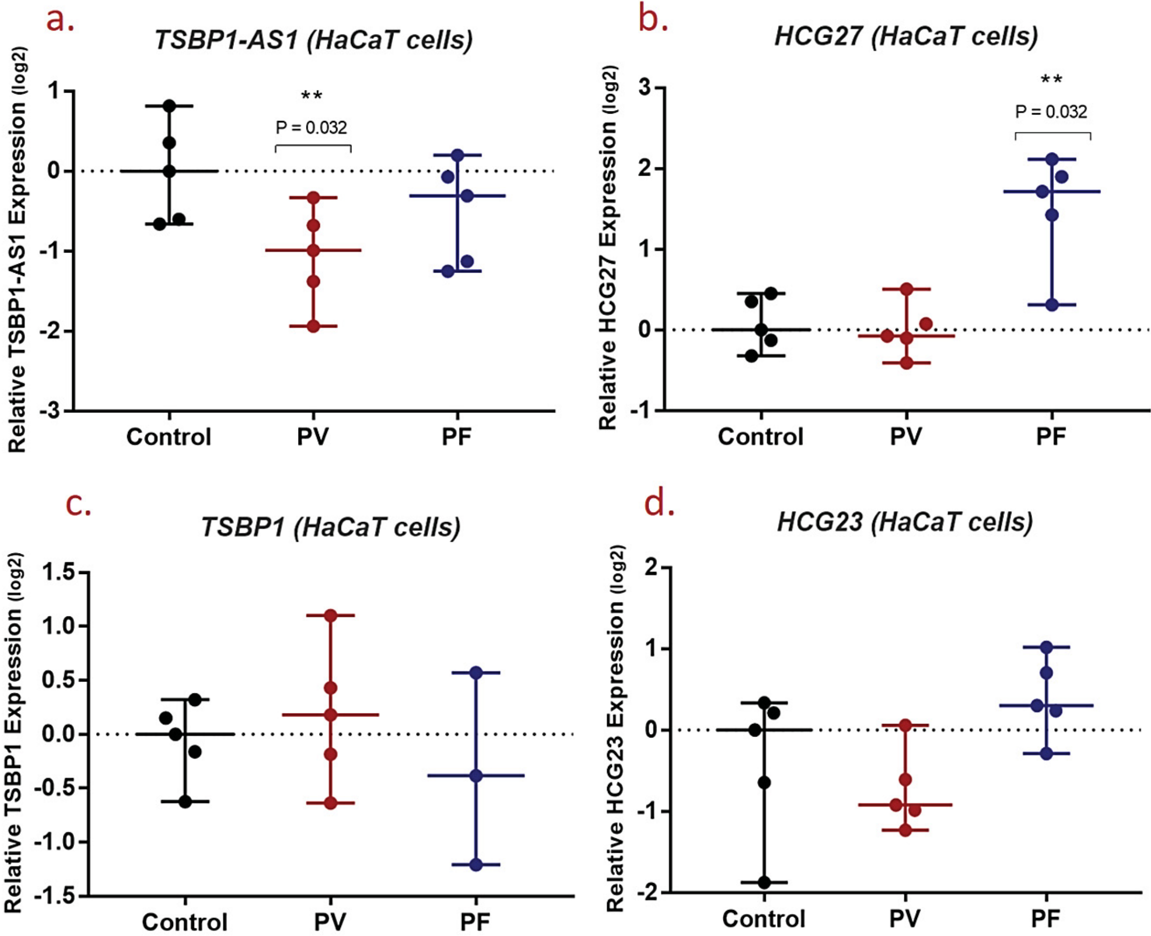

Relative expression of TSBP1-AS1 **(a)** and HCG27 **(b)** lncRNAs in PBMCs of PV patients with active disease (red) and under disease remission (blue), in comparison with controls (black). The TSBP1-AS1 (a) is downregulated in active-PV PBMC samples (p = 0.029, log2 FC = -1.29). TSBP1-AS1 lower levels are also suggestive in PBMCs of PV patients under remission (p = 0.073, log2 FC = -1) (a), where in turn, HCG27 is significantly downregulated (p = 0.048, log2 FC = -2.19). The fold-change values (FC) were calculated through the $2^{-\Delta\Delta Ct}$ method and normalized by Log2. The horizontal bars at the scatter plots indicate the median. P values were calculated with *Mann–Whitney's* test.

**Figure 3.** TSBP1-AS1, TSBP1, HCG27 and HCG23 RNA levels in PBMCs of EPF patients and controls.




Relative expression of TSBP1-AS1 **(a)**, HCG27 **(b)**, TSBP1 **(c)** and HCG23 **(d)** RNAs in PBMCs of EPF patients with active disease (red), in comparison with controls (black). The HCG27 (b) is upregulated in EPF PBMCs (p = 0.035, log2 FC = 1.299), while TSBP1 (c) is downregulated (p = 0.042, log2 FC = -2.14). Suggestive lower expression was also observed for HCG23 in EPF PBMCs (p = 0.072, log2 FC = -1.99) (d). The fold-change values (FC) were calculated through the $2^{-\Delta\Delta Ct}$ method and normalized by Log2. The horizontal bars at the scatter plots indicate the median. P values were calculated with *Mann–Whitney's* test.

**Figure 4.** TSBP1-AS1, TSBP1, HCG27 and HCG23 RNA levels in stimulated HaCaT cells.




Relative expression of TSBP1-AS1 **(a)**, HCG27 **(b)**, TSBP1 **(c)** and HCG23 **(d)** RNAs in HaCaT cells stimulated with IgG of PV (red) and PF (blue) patients, in comparison with controls (black). The TSBP1-AS1 (a) is downregulated in PV-stimulated HaCaT cells (p = 0.032, log2 FC = -0.99). HCG27 is upregulated in PF-stimulated HaCaT cells (p = 0.032, log2 FC = 1.71). The fold-change values (FC) were calculated through the $2^{-\Delta\Delta Ct}$ method and normalized by log2. The horizontal bars at the scatter plots indicate the median. P values indicate a statistical significance at the 0.05 level and were calculated with *Mann–Whitney's* test.

**Table 1.** HCG genes selected for genotyping data extraction.

| lncRNA gene | Class | Chr | Initial position | Final position |
|---|---|---|---|---|
| *HCG9* | Intergenic | 6 | 29942889 | 29946187 |
| *HCG11* | Intergenic | 6 | 26521772 | 26528881 |
| *HCG14* | Intergenic | 6 | 28864307 | 28865534 |
| *HCG15* | Intergenic | 6 | 28952526 | 28956581 |
| *HCG17/ HCG18* | Intergenic | 6 | 30201816 | 30294927 |
| *HCG20* | Intergenic | 6 | 30734602 | 30762101 |
| *HCG21* | Sense-intronic | 6 | 30913264 | 30924009 |
| *HCG22* | Intergenic | 6 | 31021227 | 31027667 |
| *TSBP1-AS1/ HCG23* | Antisense | 6 | 32222417 | 32361468 |
| *HCG24* | Intergenic | 6 | 33110860 | 33115840 |
| *HCG25* | Antisense | 6 | 33217311 | 33222766 |
| *HCG26* | Intergenic | 6 | 31439006 | 31440185 |
| *HCG27* | Intergenic | 6 | 31165537 | 31196425 |

Genomic locations of HCG lncRNA genes, according to LNCIPedia DB (hg19). The *HCG18* is located within *HCG17* gene, while *HCG23* is located within *TSBP1-AS1* gene. According to LNCipedia DB, the *TSBP1-AS1* gene and its transcripts are still classified as *HCG23* (LNCipedia DB, available in https://hg19.lncipedia.org/db/gene/HCG23). Nevertheless, in the present study, we consider the whole gene (including HCG23:6) as *TSBP1-AS1*, and the specific region transcribing *ENST00000426643.1* (TSBP1-AS1-204) as *HCG23*. Chr – chromosome. Source: LNCipedia DB and RNA Central DB.

**Table 2.** Primer pair sequences used in qPCR assays.

| RNA Target | Primer Foward (5'-> 3') | Primer Reverse (5'-> 3') | Amplicon Size (nt) | Note |
|---|---|---|---|---|
| ***TSBP1-AS1*** | CTGGCGGTCCTACTCAACAC | ATCGCTTTCTTCCCTGTGACT | 76 | Amplify the transcripts NR_136245.1, NR_136246. and NR_136244.1 (without HCG23) |
| ***TSBP1*** | TGATGCAATTCACAGCCCCT | CCAGCTGACTTGCGGTTCTC | 183 | Amplify the transcripts NM_001286475.2, NM_001286474.2 and other predicted TSBP1 mRNA isoforms |
| ***HCG23*** | CCTCTCCTTCCTGTGGCTTT | TCTGAACTTGCTCTTCTGGGC | 280 | Specific to the HCG23 transcript NR_044996.1 (ENST00000426643.1; TSBP1-AS1 - 204) |
| ***HCG27*** | CCAGGAAAGTGAAAAGAGAAGCAG | CAGCAGGAGGGATCACTAAGATTT | 76 | Amplify the transcript NR_026791.1, the main HCG27 RNA isoform (ENST00000383331.4; HCG27-201). Primers designed by Saeidi et al., 2018 (with modifications) |
| ***RPL13A*** | CTCAAGGTCGTGCGTCTGAA | GGCTGTCACTGCCTGGTACT | 93 | Endogenous control |

5’-3’ sequences of forward and reverse specific primers used for RNAs quantification. Amplicon size in nucleotides (nt). Source: PrimerBlast tool.

**Table 3.** Associated SNPs in HCG lncRNAs with PV.

| SNP | Chromosome 6 Genomic position (hg19) | Odds Ratio | L95 | U95 | FDR adjusted *p* value | Result | lncRNA genes | Overlapped coding genes |
|---|---|---|---|---|---|---|---|---|
| *rs1003879*A* | 32299592 | 4.78 | 3.601 | 6.353 | **5.24E-22** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs3115562*T* | 32233814 | 4.49 | 3.211 | 6.262 | **6.78E-14** | Associated | *TSBP1-AS1* | - |
| *rs9268267*G* | 32312495 | 4.49 | 3.211 | 6.262 | **6.78E-14** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs498422*G* | 32286761 | 4.29 | 3.088 | 5.968 | **1.74E-13** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs6910071*G* | 32282854 | 0.20 | 0.14 | 0.294 | **1.37E-12** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs570963*G* | 32289594 | 2.82 | 2.139 | 3.72 | **5.34E-09** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs2273017*G* | 32337630 | 3.95 | 2.726 | 5.716 | **7.49E-09** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs3129949*A* | 32298814 | 3.72 | 2.609 | 5.312 | **7.49E-09** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs3117137*T* | 32309911 | 3.72 | 2.609 | 5.312 | **7.49E-09** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs521828*T* | 32291643 | 2.82 | 2.06 | 3.868 | **1.74E-06** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs1265762*C* | 32321115 | 2.96 | 2.073 | 4.23 | **3.54E-05** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs3129943*G* | 32338695 | 2.50 | 1.797 | 3.484 | **7.37E-04** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs6903816*G* | 32336517 | 1.92 | 1.516 | 2.438 | **8.58E-04** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs9268303*A* | 32326165 | 1.93 | 1.516 | 2.444 | **8.82E-04** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs2050190*G* | 32339076 | 0.49 | 0.362 | 0.674 | **9.19e-05** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs969893*A* | 32316081 | 1.77 | 1.372 | 2.277 | **0.0001041** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs1055569*T* | 31440082 | 1.61 | 1.299 | 1.988 | **0.0001154** | Associated | *HCG26, HCP5* | - |
| *rs3130320*T* | 32223258 | 0.46 | 0.328 | 0.6554 | **0.0001197** | Associated | *TSBP1-AS1* | - |
| *rs35016370*A* | 31177034 | 0.60 | 0.472 | 0.766 | **0.0003116** | Associated | *HCG27* | - |
| *rs2050189*C* | 32339647 | 1.77 | 1.347 | 2.323 | **0.000324** | Associated | *TSBP1-AS1* | *TSBP1* |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *rs9268326*C* | 32330153 | 0.47 | 0.326 | 0.679 | **0.0004114** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs2517523*G* | 31026434 | 1.55 | 1.251 | 1.92 | **0.0004181** | Associated | *HCG22* | - |
| *rs547261*A* | 32282033 | 0.47 | 0.328 | 0.683 | **0.0004181** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs502626*G* | 32278266 | 0.47 | 0.328 | 0.683 | **0.0004181** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs9268132*G* | 32254654 | 0.47 | 0.329 | 0.684 | **0.0004181** | Associated | *TSBP1-AS1* | - |
| *rs539703*C* | 32288462 | 0.49 | 0.34 | 0.708 | **0.0008802** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs17481190*T* | 30743014 | 1.59 | 1.25 | 2.025 | **0.0009763** | Associated | *HCG20* | - |
| *rs2395141*G* | 32282068 | 0.39 | 0.245 | 0.651 | **0.001316** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs4279480*G, rs754388621*indel* | 32290927 | 0.40 | 0.246 | 0.654 | **0.001353** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs926070*G* | 32257566 | 0.40 | 0.248 | 0.657 | **0.001409** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs7758215*T* | 32345131 | 0.54 | 0.381 | 0.755 | **0.001835** | Associated | *TSBP1-AS1* | - |
| *rs916571*A* | 30201955 | 0.68 | 0.554 | 0.845 | **0.002158** | Associated | *HCG17* | - |
| *rs3129927*C* | 32333827 | 0.26 | 0.126 | 0.556 | **0.00219** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs6930777*T* | 32351566 | 0.49 | 0.327 | 0.731 | **0.002246** | Associated | *TSBP1-AS1* | - |
| *rs1265757*T* | 32302382 | 0.27 | 0.127 | 0.563 | **0.002246** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs7775397*G, rs1379518282*indel* | 32261252 | 0.27 | 0.127 | 0.563 | **0.002246** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs2395149*A* | 32325562 | 0.27 | 0.127 | 0.564 | **0.002246** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs9268177*A* | 32274882 | 0.27 | 0.128 | 0.567 | **0.002264** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs3132971*G* | 32230256 | 0.27 | 0.128 | 0.567 | **0.002264** | Associated | *TSBP1-AS1* | - |
| *rs2285802*C* | 30209099 | 0.69 | 0.562 | 0.854 | **0.002319** | Associated | *HCG17* | - |
| *rs2523857*T* | 31021504 | 0.68 | 0.549 | 0.85 | **0.002522** | Associated | *HCG22* | - |
| *rs17191181*T* | 31171876 | 0.61 | 0.456 | 0.81 | **0.002611** | Associated | *HCG27* | - |
| *rs28397284*C* | 31166936 | 0.62 | 0.461 | 0.82 | **0.003486** | Associated | *HCG27* | - |
| *rs17208671*T* | 32360849 | 0.60 | 0.445 | 0.816 | **0.003809** | Associated | *TSBP1-AS1, HCG23* | - |
| *rs9268219*G* | 32284108 | 0.29 | 0.138 | 0.609 | **0.003823** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs3117098*G, rs938761994*indel* | 32358513 | 1.47 | 1.166 | 1.862 | **0.004095** | Associated | *TSBP1-AS1, HCG23* | - |
| *rs3129924*T* | 32333299 | 0.36 | 0.189 | 0.678 | **0.005483** | Associated | *TSBP1-AS1* | *TSBP1* |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *rs28744251 *T* | 31177874 | 0.69 | 0.543 | 0.87 | **0.00593** | Associated | *HCG27* | - |
| *rs9368628*G* | 30238559 | 1.46 | 1.15 | 1.857 | **0.006312** | Associated | *HCG17* | - |
| *rs1579219*T* | 30224305 | 1.46 | 1.147 | 1.846 | **0.006412** | Associated | *HCG17* | - |
| *rs9366755*T* | 30286729 | 1.45 | 1.144 | 1.841 | **0.006741** | Associated | *HCG17, HCG18* | - |
| *rs9262620*A* | 31022489 | 1.44 | 1.14 | 1.814 | **0.006741** | Associated | *HCG22* | - |
| *rs3131043*G* | 30758466 | 1.40 | 1.128 | 1.743 | **0.006939** | Associated | *HCG20* | - |
| *rs2516697*A* | 30232514 | 1.42 | 1.131 | 1.778 | **0.007412** | Associated | *HCG17* | *HLA-L* |
| *rs28362351*T* | 31167760 | 1.54 | 1.155 | 2.063 | **0.009848** | Associated | *HCG27* | - |
| *rs885915*T* | 30202715 | 0.71 | 0.564 | 0.897 | 0.01138 | Suggestive | *HCG17* | - |
| *rs9391858*G* | 32341398 | 0.63 | 0.453 | 0.861 | 0.01138 | Suggestive | *TSBP1-AS1* | - |
| *rs72863820*A* | 31182682 | 1.56 | 1.15 | 2.12 | 0.01181 | Suggestive | *HCG27* | - |
| *rs3096683*G* | 32234993 | 0.39 | 0.203 | 0.746 | 0.01196 | Suggestive | *TSBP1-AS1* | - |
| *rs9268127*C* | 32253559 | 0.39 | 0.205 | 0.749 | 0.01196 | Suggestive | *TSBP1-AS1* | - |
| *rs1033499*A* | 32307532 | 0.39 | 0.205 | 0.749 | 0.01196 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs9268220*T* | 32284340 | 0.39 | 0.205 | 0.749 | 0.01196 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs6910668*G* | 32263458 | 0.39 | 0.205 | 0.751 | 0.01211 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs9380167*G* | 30274461 | 1.36 | 1.094 | 1.687 | 0.01397 | Suggestive | *HCG17, HCG18* | - |
| *rs12662501*T* | 31190850 | 1.41 | 1.09 | 1.828 | 0.022 | Suggestive | *HCG27* | - |
| *rs3117119*G* | 32318610 | 0.52 | 0.313 | 0.853 | 0.02376 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs3130667*A* | 30743241 | 0.74 | 0.591 | 0.936 | 0.02736 | Suggestive | *HCG20* | - |
| *rs3129945*A* | 32342537 | 0.65 | 0.468 | 0.909 | 0.02766 | Suggestive | *TSBP1-AS1* | - |
| *rs9268205*A* | 32279938 | 0.53 | 0.322 | 0.869 | 0.02784 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs9268234*A* | 32289390 | 0.53 | 0.323 | 0.873 | 0.0287 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs16898922*T* | 31171224 | 1.57 | 1.093 | 2.241 | 0.03236 | Suggestive | *HCG27* | - |
| *rs3132958*A* | 32297901 | 0.55 | 0.341 | 0.894 | 0.03412 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs3132944*T* | 32307446 | 0.55 | 0.341 | 0.894 | 0.03412 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs3117128*A* | 32315500 | 0.55 | 0.341 | 0.894 | 0.03412 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs3129931*C* | 32335516 | 0.55 | 0.342 | 0.896 | 0.03447 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs9268103*A* | 32245370 | 0.55 | 0.332 | 0.899 | 0.03635 | Suggestive | *TSBP1-AS1* | - |
| *rs3131788*A* | 31024796 | 0.50 | 0.279 | 0.894 | 0.04028 | Suggestive | *HCG22* | - |

In bold: PV-associated SNPs with FDR-adjusted $p < 0.05$. Suggestive associations: FDR-adjusted p values between 0.05 and 0.1.

L95 - lower limit of 95% confidence interval; U95 - upper limit of 95% confidence interval.

**Table 4.** Associated SNPs in HCG lncRNAs with EPF.

| SNP | Chromosome 6 Genomic position (hg19) | Odds Ratio | L95 | U95 | FDR-adjusted *p* value | Result | lncRNA genes | Overlapped coding genes |
|---|---|---|---|---|---|---|---|---|
| *rs16870005*A* | 32261153 | 10.56 | 4.415 | 25.26 | **0.01351** | Associated | *TSBP1-AS1* | *TSBP1* |
| *rs3129900*C* | 32305979 | 1.97 | 1.353 | 2.878 | **0.02176** | Associated | *TSBP1-AS1* | *TSBP1* |
| <u>*rs3129943*G*</u> | 32338695 | 0.53 | 0.363 | 0.769 | **0.02176** | Associated | *TSBP1-AS1* | *TSBP1* |
| <u>*rs3129949*A*</u> | 32298814 | 0.52 | 0.351 | 0.766 | **0.02176** | Associated | *TSBP1-AS1* | *TSBP1* |
| <u>*rs9268103*A*</u> | 32245370 | 1.87 | 1.284 | 2.715 | **0.02176** | Associated | *TSBP1-AS1* | |
| *rs1634703*A* | 31185262 | 2.78 | 1.501 | 5.136 | **0.02176** | Associated | *HCG27:6* | |
| *rs79792575*A* | 30920086 | 2.58 | 1.444 | 4.612 | **0.02261** | Associated | *HCG21* | *MUCL3, SFTA2* |
| *rs3132931*C* | 32235895 | 0.55 | 0.371 | 0.814 | **0.04028** | Associated | *TSBP1-AS1* | |
| *rs9468829*C* | 30749233 | 1.76 | 1.204 | 2.57 | **0.04489** | Associated | *HCG20* | |
| *rs9268168*A* | 32272510 | 0.57 | 0.39 | 0.842 | 0.05224 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs3129945*A* | 32342537 | 1.56 | 1.144 | 2.128 | 0.05224 | Suggestive | *TSBP1-AS1* | |
| *rs3130521*A* | 31196376 | 0.68 | 0.503 | 0.907 | 0.08405 | Suggestive | *HCG27* | |
| *rs9268220*A* | 32284340 | 0.51 | 0.301 | 0.846 | 0.08405 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs6910668*C* | 32263458 | 0.51 | 0.304 | 0.856 | 0.08839 | Suggestive | *TSBP1-AS1* | *TSBP1* |
| *rs11757629*C* | 30744529 | 0.47 | 0.257 | 0.852 | 0.0994 | Suggestive | *HCG20* | |

SNPs associated with EPF, by logistic regression. The underlined SNPs are also associated with PV. In bold: PF-associated SNPs with FDR-adjusted $p < 0.05$.

Suggestive associations: FDR-adjusted p values between 0.05 and 0.1. L95 - lower limit of 95% confidence interval; U95 - upper limit of 95% confidence interval.

**Supplementary material**

**Suppl. Table 1.** Clinical and demographic aspects of subjects.

| Group | Cohort | N | % female | Age mean | Samples used for |
|---|---|---|---|---|---|
| EPF patients | Brazilian endemic areas | 227 | 50.6 | 41.4 | EPF Microarray |
| EPF controls | | 194 | 51.3 | 44.8 | |
| PV patients | Germany | 241 | 47.6 | n.i. | PV Microarray |
| PV controls | | 1188 | | n.i. | |
| EPF patients | Brazilian endemic areas | 6 (100% with active disease) | 50 | 54.3 | qPCRs (of PBMCs) |
| PV patients | Brazilian endemic and non-endemic areas | 15 (54% with active disease) | 73.3 | 40.3 | |
| Controls | | 7 (28.6% endemic) | 71.4 | 45.6 | |
| PF patients | Germany | 5 | n.i. | 50.4 | Isolation of IgG antibodies used in cell culture; qPCRs |
| PV patients | | 5 | n.i. | 52.2 | |
| Controls | | 6 | 50 | 58.8 | |

Demographic and clinical data of subjects cohorts enrolled in this study. N – number of subjects; n.i. – non-informed data.

**Supp. Table 2.** Genomic positions of the analyzed SNPs at HCG genes.

| | Pemphigus vulgaris | | | Endemic pemphigus foliaceus | | |
|---|---|---|---|---|---|---|
| **Chr** | **Position (hg19)** | **A1** | **A2** | **Position (hg19)** | **A1** | **A2** |
| 6 | 26523759 | A | C | 26510748 | A | G |
| 6 | 28955526 | G | A | 28918936 | A | G |
| 6 | 29942927 | G | T | 29943035 | A | G |
| 6 | 29943035 | A | G | 29943067 | A | G |
| 6 | 29943281 | A | G | 29943832 | G | A |
| 6 | 29943490 | A | G | 30202571 | A | G |
| 6 | 29943528 | A | G | 30209062 | C | A |
| 6 | 29943581 | A | C | 30210807 | A | C |
| 6 | 29943656 | T | G | 30211755 | A | G |
| 6 | 29943715 | T | C | 30211903 | C | G |
| 6 | 29943832 | C | T | 30222256 | A | G |
| 6 | 29943981 | A | G | 30224305 | A | G |
| 6 | 29944004 | C | A | 30224889 | A | G |
| 6 | 29944148 | A | C | 30228138 | A | G |
| 6 | 29944158 | C | T | 30233558 | A | C |
| 6 | 29944184 | T | C | 30252836 | A | T |
| 6 | 29945368 | G | A | 30257846 | T | A |
| 6 | 29945508 | T | C | 30262512 | G | A |
| 6 | 29945602 | A | G | 30269561 | A | T |
| 6 | 29945620 | T | C | 30272417 | C | G |
| 6 | 29945741 | T | C | 30275246 | C | A |
| 6 | 29945771 | T | G | 30281234 | A | T |
| 6 | 29945841 | A | G | 30282332 | A | G |
| 6 | 29945898 | C | T | 30285650 | T | A |
| 6 | 29945949 | A | G | 30286729 | A | G |
| 6 | 30201955 | A | G | 30289271 | A | G |
| 6 | 30202429 | T | C | 30293483 | A | G |
| 6 | 30202571 | T | C | 30735105 | G | A |
| 6 | 30202715 | T | C | 30735229 | A | G |
| 6 | 30204526 | T | G | 30738476 | A | G |
| 6 | 30205407 | T | G | 30739846 | A | G |
| 6 | 30205991 | C | T | 30739904 | A | G |
| 6 | 30206014 | A | G | 30742134 | A | G |
| 6 | 30206186 | A | G | 30743241 | A | G |
| 6 | 30206726 | G | T | 30744529 | C | A |
| 6 | 30207495 | A | C | 30749233 | C | A |
| 6 | 30209062 | G | T | 30760025 | G | A |
| 6 | 30209099 | C | T | 30760698 | C | A |
| 6 | 30209493 | G | A | 30761734 | G | A |
| 6 | 30209802 | G | T | 30914751 | G | A |
| 6 | 30209836 | C | T | 30914843 | A | G |
| 6 | 30209918 | G | A | 30919701 | G | A |

| | | | | | | |
|---|---|---|---|---|---|---|
| 6 | 30209977 | T | G | 30919878 | C | G |
| 6 | 30210407 | C | T | 30920086 | A | G |
| 6 | 30211001 | G | A | 30920957 | A | G |
| 6 | 30211660 | G | A | 30921882 | G | A |
| 6 | 30211755 | T | C | 31022113 | C | G |
| 6 | 30211805 | A | G | 31024808 | G | A |
| 6 | 30211894 | T | G | 31025051 | G | A |
| 6 | 30212504 | T | C | 31025989 | A | G |
| 6 | 30213924 | T | C | 31026434 | A | G |
| 6 | 30214347 | A | G | 31027516 | A | G |
| 6 | 30214751 | A | G | 31167927 | A | G |
| 6 | 30215367 | G | A | 31170528 | G | A |
| 6 | 30215592 | A | G | 31174527 | A | G |
| 6 | 30221921 | C | T | 31176226 | G | A |
| 6 | 30222020 | T | C | 31176335 | A | G |
| 6 | 30222256 | A | G | 31176921 | A | G |
| 6 | 30223903 | T | C | 31177915 | G | A |
| 6 | 30224238 | T | G | 31184196 | A | G |
| 6 | 30224305 | T | C | 31185262 | A | G |
| 6 | 30224889 | T | C | 31190303 | G | A |
| 6 | 30225845 | C | T | 31190850 | A | G |
| 6 | 30227206 | A | G | 31192796 | G | A |
| 6 | 30227397 | T | C | 31193155 | G | A |
| 6 | 30228138 | A | G | 31195218 | A | G |
| 6 | 30228721 | A | G | 31196376 | A | G |
| 6 | 30229306 | G | A | 31439063 | G | A |
| 6 | 30229869 | T | C | 31439740 | C | A |
| 6 | 30229989 | T | C | 31440082 | A | G |
| 6 | 30230661 | C | T | 32223258 | A | G |
| 6 | 30230760 | A | G | 32235757 | G | A |
| 6 | 30230930 | G | A | 32255269 | A | G |
| 6 | 30231273 | G | A | 32257566 | G | A |
| 6 | 30231330 | C | T | 32259527 | G | A |
| 6 | 30231587 | G | A | 32261153 | A | G |
| 6 | 30231666 | C | T | 32261291 | G | C |
| 6 | 30231768 | A | G | 32263458 | C | A |
| 6 | 30232009 | G | T | 32266425 | A | G |
| 6 | 30232374 | A | G | 32268701 | C | A |
| 6 | 30232436 | A | G | 32282854 | G | A |
| 6 | 30232514 | A | G | 32282979 | G | A |
| 6 | 30232672 | T | C | 32283844 | G | A |
| 6 | 30232953 | A | G | 32284340 | A | G |
| 6 | 30233558 | A | C | 32286761 | C | A |
| 6 | 30234152 | T | C | 32291359 | A | C |
| 6 | 30234494 | C | T | 32299822 | A | G |
| 6 | 30234657 | G | A | 32305979 | C | A |

| | | | |
|---|---|---|---|
| 6 | 30234721 | A | G |
| 6 | 30235046 | T | G |
| 6 | 30235184 | T | C |
| 6 | 30235204 | A | G |
| 6 | 30236754 | C | T |
| 6 | 30237454 | G | A |
| 6 | 30238559 | G | A |
| 6 | 30238581 | G | A |
| 6 | 30241114 | G | A |
| 6 | 30242128 | G | A |
| 6 | 30243018 | C | T |
| 6 | 30243439 | A | G |
| 6 | 30243947 | G | A |
| 6 | 30245033 | A | G |
| 6 | 30246212 | A | C |
| 6 | 30256838 | A | G |
| 6 | 30256885 | T | C |
| 6 | 30256936 | T | C |
| 6 | 30257693 | C | T |
| 6 | 30257751 | C | T |
| 6 | 30257781 | G | A |
| 6 | 30258708 | T | C |
| 6 | 30259272 | T | C |
| 6 | 30259657 | T | G |
| 6 | 30260982 | T | C |
| 6 | 30261032 | T | G |
| 6 | 30262082 | C | T |
| 6 | 30262512 | G | A |
| 6 | 30263566 | G | T |
| 6 | 30264307 | T | C |
| 6 | 30264624 | T | C |
| 6 | 30269923 | T | C |
| 6 | 30270080 | A | G |
| 6 | 30270270 | A | G |
| 6 | 30270609 | T | C |
| 6 | 30270684 | C | T |
| 6 | 30271334 | G | A |
| 6 | 30271792 | A | G |
| 6 | 30272221 | C | T |
| 6 | 30273939 | T | C |
| 6 | 30274461 | G | A |
| 6 | 30274773 | T | G |
| 6 | 30275124 | G | T |
| 6 | 30278112 | T | C |
| 6 | 30279059 | T | C |
| 6 | 30279130 | A | G |
| | 32315654 | C | G |
| | 32317276 | G | A |
| | 32321004 | A | G |
| | 32321597 | G | A |
| | 32331998 | A | G |
| | 32333195 | A | G |
| | 32335204 | A | G |
| | 32335433 | A | G |
| | 32336586 | G | A |
| | 32337630 | G | A |
| | 32337686 | A | G |
| | 32338695 | G | A |
| | 32338986 | A | G |
| | 32339076 | G | A |
| | 32339348 | G | A |
| | 32339647 | G | A |
| | 32339840 | A | T |
| | 32342537 | A | G |
| | 32345595 | C | G |
| | 32351566 | A | G |
| | 32357165 | G | A |
| | 32358270 | A | G |
| | 32358513 | G | A |
| | 32361111 | A | G |
| | 32361388 | A | G |
| | 33111347 | A | G |
| | 33114171 | A | G |
| | 33115024 | G | A |
| | 33222163 | A | C |

| | | | |
|---|---|---|---|
| 6 | 30280125 | C | A |
| 6 | 30281336 | T | G |
| 6 | 30281560 | C | T |
| 6 | 30281578 | A | G |
| 6 | 30282332 | T | C |
| 6 | 30284457 | A | G |
| 6 | 30284651 | T | C |
| 6 | 30284920 | C | A |
| 6 | 30285121 | T | C |
| 6 | 30285405 | A | C |
| 6 | 30285524 | T | C |
| 6 | 30285943 | C | T |
| 6 | 30286729 | T | C |
| 6 | 30287085 | C | T |
| 6 | 30288924 | C | T |
| 6 | 30289706 | T | C |
| 6 | 30292083 | C | T |
| 6 | 30292147 | T | C |
| 6 | 30293186 | A | G |
| 6 | 30293483 | T | C |
| 6 | 30293585 | T | G |
| 6 | 30293592 | T | C |
| 6 | 30294590 | A | G |
| 6 | 30735105 | G | A |
| 6 | 30735229 | A | G |
| 6 | 30736010 | T | C |
| 6 | 30736360 | T | G |
| 6 | 30737486 | C | A |
| 6 | 30738042 | T | C |
| 6 | 30738211 | T | C |
| 6 | 30738446 | A | G |
| 6 | 30738468 | A | G |
| 6 | 30738600 | G | A |
| 6 | 30738718 | T | C |
| 6 | 30739904 | A | G |
| 6 | 30740038 | A | G |
| 6 | 30740160 | A | G |
| 6 | 30740256 | A | G |
| 6 | 30740515 | A | C |
| 6 | 30742141 | A | G |
| 6 | 30743014 | T | C |
| 6 | 30743241 | A | G |
| 6 | 30744529 | C | A |
| 6 | 30745454 | C | T |
| 6 | 30745489 | A | C |
| 6 | 30746498 | T | C |

| 6 | 30746519 | T | G |
|---|---|---|---|
| 6 | 30746737 | A | G |
| 6 | 30748297 | G | A |
| 6 | 30749233 | G | T |
| 6 | 30749589 | A | G |
| 6 | 30755933 | A | G |
| 6 | 30758466 | G | A |
| 6 | 30758487 | G | A |
| 6 | 30758555 | T | C |
| 6 | 30759188 | G | T |
| 6 | 30759283 | A | G |
| 6 | 30760025 | C | T |
| 6 | 30760698 | T | G |
| 6 | 30761734 | C | T |
| 6 | 30913458 | A | G |
| 6 | 30914552 | C | T |
| 6 | 30914751 | C | T |
| 6 | 30914843 | T | C |
| 6 | 30915455 | A | G |
| 6 | 30916259 | A | G |
| 6 | 30917482 | T | C |
| 6 | 30919391 | A | G |
| 6 | 30919701 | C | T |
| 6 | 30920086 | T | C |
| 6 | 30920124 | A | G |
| 6 | 30920890 | A | G |
| 6 | 30921141 | C | T |
| 6 | 30922532 | T | C |
| 6 | 31021504 | T | C |
| 6 | 31021547 | A | G |
| 6 | 31022169 | G | A |
| 6 | 31022354 | G | A |
| 6 | 31022489 | A | G |
| 6 | 31023203 | G | T |
| 6 | 31023840 | G | A |
| 6 | 31024796 | A | G |
| 6 | 31024808 | G | A |
| 6 | 31025848 | G | A |
| 6 | 31026236 | G | T |
| 6 | 31026434 | G | A |
| 6 | 31027031 | T | C |
| 6 | 31027516 | A | G |
| 6 | 31027569 | A | G |
| 6 | 31165566 | T | C |
| 6 | 31166072 | A | C |
| 6 | 31166887 | T | G |

| | | | |
|---|---|---|---|
| 6 | 31166936 | C | T |
| 6 | 31167498 | C | T |
| 6 | 31167512 | G | A |
| 6 | 31167573 | A | G |
| 6 | 31167760 | T | C |
| 6 | 31167852 | G | A |
| 6 | 31167927 | T | C |
| 6 | 31168029 | A | G |
| 6 | 31168397 | A | G |
| 6 | 31168483 | G | A |
| 6 | 31168494 | A | G |
| 6 | 31169289 | C | T |
| 6 | 31170014 | C | T |
| 6 | 31170514 | G | A |
| 6 | 31170608 | T | C |
| 6 | 31170713 | C | T |
| 6 | 31170914 | G | A |
| 6 | 31170947 | C | T |
| 6 | 31171224 | T | G |
| 6 | 31171257 | C | A |
| 6 | 31171275 | A | G |
| 6 | 31171876 | T | C |
| 6 | 31171924 | G | A |
| 6 | 31172151 | C | T |
| 6 | 31172655 | A | G |
| 6 | 31172795 | A | G |
| 6 | 31173806 | C | T |
| 6 | 31174521 | T | C |
| 6 | 31174527 | G | A |
| 6 | 31176226 | C | T |
| 6 | 31176921 | A | G |
| 6 | 31177034 | A | G |
| 6 | 31177094 | A | G |
| 6 | 31177874 | T | C |
| 6 | 31177915 | G | A |
| 6 | 31178334 | G | A |
| 6 | 31178901 | A | G |
| 6 | 31182383 | G | A |
| 6 | 31182682 | A | C |
| 6 | 31182833 | G | T |
| 6 | 31184175 | A | G |
| 6 | 31184418 | T | C |
| 6 | 31184425 | A | G |
| 6 | 31185770 | A | G |
| 6 | 31185810 | G | T |
| 6 | 31186039 | C | T |

| | | | |
|---|---|---|---|
| 6 | 31186208 | G | A |
| 6 | 31186230 | T | C |
| 6 | 31186245 | A | G |
| 6 | 31186378 | A | G |
| 6 | 31187075 | T | C |
| 6 | 31188267 | T | C |
| 6 | 31188312 | C | T |
| 6 | 31188315 | C | A |
| 6 | 31189219 | A | G |
| 6 | 31189543 | G | A |
| 6 | 31190303 | C | T |
| 6 | 31190850 | T | C |
| 6 | 31190977 | A | C |
| 6 | 31191509 | T | C |
| 6 | 31192766 | T | C |
| 6 | 31192842 | T | C |
| 6 | 31192977 | C | T |
| 6 | 31194690 | T | C |
| 6 | 31194747 | G | A |
| 6 | 31195218 | T | C |
| 6 | 31195269 | A | G |
| 6 | 31195365 | G | T |
| 6 | 31195485 | T | C |
| 6 | 31195616 | A | G |
| 6 | 31196376 | T | C |
| 6 | 31440082 | T | C |
| 6 | 32223258 | T | C |
| 6 | 32224139 | A | G |
| 6 | 32224463 | C | T |
| 6 | 32230256 | G | T |
| 6 | 32230608 | C | T |
| 6 | 32233814 | T | C |
| 6 | 32234015 | A | G |
| 6 | 32234922 | C | T |
| 6 | 32234993 | G | A |
| 6 | 32235177 | G | A |
| 6 | 32235757 | G | A |
| 6 | 32235895 | G | T |
| 6 | 32236054 | T | G |
| 6 | 32237926 | T | C |
| 6 | 32238013 | C | T |
| 6 | 32238219 | A | G |
| 6 | 32239651 | T | C |
| 6 | 32244627 | C | T |
| 6 | 32245370 | A | G |
| 6 | 32245827 | T | C |

| | | | |
|---|---|---|---|
| 6 | 32245836 | T | C |
| 6 | 32252678 | C | T |
| 6 | 32253559 | C | T |
| 6 | 32254452 | G | A |
| 6 | 32254654 | G | A |
| 6 | 32255240 | G | A |
| 6 | 32255269 | A | G |
| 6 | 32255417 | A | G |
| 6 | 32255507 | G | A |
| 6 | 32257337 | G | A |
| 6 | 32257566 | G | A |
| 6 | 32259527 | G | A |
| 6 | 32260350 | C | T |
| 6 | 32260559 | G | A |
| 6 | 32261153 | T | C |
| 6 | 32261252 | G | T |
| 6 | 32261507 | C | T |
| 6 | 32261630 | A | G |
| 6 | 32261771 | G | A |
| 6 | 32261952 | A | G |
| 6 | 32263458 | G | T |
| 6 | 32265528 | T | C |
| 6 | 32265881 | T | G |
| 6 | 32266425 | T | C |
| 6 | 32266490 | C | T |
| 6 | 32266795 | A | G |
| 6 | 32268701 | G | T |
| 6 | 32269578 | C | T |
| 6 | 32270500 | G | A |
| 6 | 32271807 | T | G |
| 6 | 32272327 | G | A |
| 6 | 32272433 | C | T |
| 6 | 32272437 | A | C |
| 6 | 32273765 | G | A |
| 6 | 32274882 | A | C |
| 6 | 32275194 | A | G |
| 6 | 32277934 | A | C |
| 6 | 32278187 | T | C |
| 6 | 32278266 | G | A |
| 6 | 32278635 | G | A |
| 6 | 32278792 | T | C |
| 6 | 32279340 | T | C |
| 6 | 32279532 | A | C |
| 6 | 32279622 | T | G |
| 6 | 32279816 | A | G |
| 6 | 32279938 | A | G |

| 6 | 32280182 | G | A |
|---|---|---|---|
| 6 | 32281360 | C | T |
| 6 | 32281688 | T | C |
| 6 | 32281889 | C | T |
| 6 | 32282033 | A | G |
| 6 | 32282068 | G | A |
| 6 | 32282081 | G | A |
| 6 | 32282471 | G | A |
| 6 | 32282854 | G | A |
| 6 | 32282979 | G | A |
| 6 | 32283494 | A | G |
| 6 | 32283844 | C | T |
| 6 | 32284108 | G | T |
| 6 | 32284340 | T | C |
| 6 | 32285596 | A | G |
| 6 | 32286761 | G | T |
| 6 | 32288089 | C | T |
| 6 | 32288238 | G | T |
| 6 | 32288462 | C | A |
| 6 | 32289240 | T | C |
| 6 | 32289318 | C | T |
| 6 | 32289390 | A | C |
| 6 | 32289594 | G | A |
| 6 | 32290208 | T | C |
| 6 | 32290927 | G | T |
| 6 | 32290954 | G | A |
| 6 | 32291190 | G | A |
| 6 | 32291643 | T | C |
| 6 | 32291690 | G | A |
| 6 | 32291837 | T | G |
| 6 | 32292084 | T | C |
| 6 | 32292715 | A | G |
| 6 | 32292956 | T | C |
| 6 | 32294843 | C | T |
| 6 | 32294992 | A | G |
| 6 | 32295350 | A | G |
| 6 | 32295357 | A | G |
| 6 | 32296780 | T | C |
| 6 | 32297209 | A | G |
| 6 | 32297337 | C | T |
| 6 | 32297901 | A | G |
| 6 | 32297906 | T | G |
| 6 | 32298372 | G | A |
| 6 | 32298814 | A | C |
| 6 | 32298942 | G | A |
| 6 | 32299592 | A | G |

| | | | |
|---|---|---|---|
| 6 | 32299822 | A | G |
| 6 | 32299873 | T | G |
| 6 | 32301289 | C | T |
| 6 | 32301322 | T | C |
| 6 | 32301514 | A | G |
| 6 | 32302382 | T | C |
| 6 | 32303068 | A | G |
| 6 | 32303474 | T | G |
| 6 | 32303511 | A | G |
| 6 | 32304031 | A | G |
| 6 | 32305370 | G | T |
| 6 | 32306090 | A | G |
| 6 | 32306559 | T | C |
| 6 | 32306922 | C | T |
| 6 | 32307137 | C | T |
| 6 | 32307260 | G | T |
| 6 | 32307382 | A | G |
| 6 | 32307446 | T | C |
| 6 | 32307532 | A | G |
| 6 | 32308908 | A | G |
| 6 | 32309280 | G | A |
| 6 | 32309323 | T | C |
| 6 | 32309352 | C | T |
| 6 | 32309911 | T | C |
| 6 | 32310967 | T | C |
| 6 | 32312495 | G | A |
| 6 | 32313143 | C | A |
| 6 | 32313461 | T | G |
| 6 | 32313659 | A | G |
| 6 | 32314676 | A | G |
| 6 | 32315173 | C | T |
| 6 | 32315486 | G | A |
| 6 | 32315500 | A | G |
| 6 | 32315727 | T | C |
| 6 | 32315929 | A | G |
| 6 | 32316081 | A | C |
| 6 | 32316604 | T | C |
| 6 | 32317131 | G | A |
| 6 | 32317215 | C | T |
| 6 | 32317276 | C | T |
| 6 | 32317471 | C | A |
| 6 | 32317635 | A | G |
| 6 | 32317973 | G | A |
| 6 | 32318036 | G | A |
| 6 | 32318610 | G | A |
| 6 | 32318644 | C | T |

| | | | |
|---|---|---|---|
| 6 | 32319637 | A | G |
| 6 | 32320153 | G | A |
| 6 | 32321115 | C | A |
| 6 | 32321272 | C | A |
| 6 | 32321554 | T | G |
| 6 | 32322393 | C | T |
| 6 | 32323529 | A | G |
| 6 | 32323745 | G | A |
| 6 | 32325292 | A | C |
| 6 | 32325358 | T | C |
| 6 | 32325562 | A | G |
| 6 | 32325658 | A | G |
| 6 | 32325707 | C | A |
| 6 | 32326045 | G | A |
| 6 | 32326165 | A | G |
| 6 | 32327781 | C | T |
| 6 | 32330153 | C | T |
| 6 | 32330695 | G | A |
| 6 | 32331841 | A | G |
| 6 | 32333093 | A | G |
| 6 | 32333098 | C | T |
| 6 | 32333195 | A | G |
| 6 | 32333299 | T | C |
| 6 | 32333480 | C | T |
| 6 | 32333827 | C | A |
| 6 | 32333955 | C | T |
| 6 | 32334754 | T | C |
| 6 | 32335013 | A | G |
| 6 | 32335204 | A | G |
| 6 | 32335347 | A | G |
| 6 | 32335433 | A | G |
| 6 | 32335516 | C | T |
| 6 | 32336074 | T | C |
| 6 | 32336161 | A | G |
| 6 | 32336517 | G | A |
| 6 | 32336586 | G | A |
| 6 | 32336766 | G | A |
| 6 | 32337630 | G | A |
| 6 | 32338695 | G | A |
| 6 | 32338986 | T | C |
| 6 | 32339076 | G | A |
| 6 | 32339647 | C | T |
| 6 | 32341318 | G | A |
| 6 | 32341353 | A | G |
| 6 | 32341398 | G | A |
| 6 | 32341473 | C | T |

| | | | |
|---|---|---|---|
| 6 | 32341931 | T | C |
| 6 | 32342537 | A | G |
| 6 | 32344973 | T | G |
| 6 | 32345131 | T | G |
| 6 | 32345183 | G | A |
| 6 | 32345283 | A | C |
| 6 | 32345891 | G | T |
| 6 | 32347490 | T | C |
| 6 | 32349737 | G | T |
| 6 | 32349946 | A | C |
| 6 | 32350454 | A | G |
| 6 | 32351566 | T | C |
| 6 | 32354428 | G | A |
| 6 | 32355109 | G | A |
| 6 | 32355605 | A | G |
| 6 | 32355808 | A | G |
| 6 | 32356097 | G | A |
| 6 | 32356272 | C | T |
| 6 | 32356917 | A | G |
| 6 | 32357023 | A | G |
| 6 | 32357133 | T | C |
| 6 | 32357165 | C | T |
| 6 | 32357274 | T | C |
| 6 | 32357333 | G | A |
| 6 | 32357715 | T | C |
| 6 | 32358113 | T | C |
| 6 | 32358231 | T | C |
| 6 | 32358368 | C | T |
| 6 | 32358459 | T | G |
| 6 | 32358513 | G | A |
| 6 | 32358533 | G | A |
| 6 | 32358883 | C | T |
| 6 | 32359389 | A | G |
| 6 | 32359460 | A | G |
| 6 | 32359821 | A | G |
| 6 | 32360589 | G | A |
| 6 | 32360644 | T | C |
| 6 | 32360849 | T | G |
| 6 | 32360915 | G | A |
| 6 | 32360955 | A | G |
| 6 | 32361003 | T | C |
| 6 | 32361080 | G | T |
| 6 | 32361251 | A | G |
| 6 | 32361388 | T | C |
| 6 | 33110933 | G | A |
| 6 | 33111347 | T | C |

| | | | |
|---|---|---|---|
| 6 | 33112601 | T | G |
| 6 | 33112640 | T | G |
| 6 | 33113197 | T | C |
| 6 | 33114894 | T | G |
| 6 | 33115024 | C | T |
| 6 | 33115062 | T | C |

**Supp. Table 3.** Functional annotations for the most associated SNPs.

| SNP | Alleles | MAF | Gene region | Linkage disequilibrium | Regulatory and eQTL features | Disease associated in this study |
|---|---|---|---|---|---|---|
| *rs16870005*A* | G>A | 1% EUR/AFR; 3% AMR | Intron 4 TSBP1-AS1; exon 30 TSBP1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 and neighbour genes | - | EPF associated |
| *rs3129900*C* | A>C | 14% EUR; 22% AFR; 12% AMR | Intron 4 TSBP1-AS1; intron 14 TSBP1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region and nearby | eQTL for HCG23, HLA-DRB1, -DRB5, -DRB6, -DQB1, C4A and others in skin tissues; sQTL for HLA-DRB1, -DRB5, -DRB6, -DQB1, -DQB2 and others in skin tissues | EPF associated |
| *rs3129943*G* | A>G | 25% EUR; 32% AFR; 17% AMR | Intron 6 TSBP1-AS1; intron 1 TSBP1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region | TF binding sites for SOX10, SOX8, SOX9; GMEB2; promoter flanking region; eQTL for HLA-DQB1, -DRB5, C4A, C4B and others in skin tissues; sQTL for HLA-DQB1, -DQB2 and -DPB2 in skin tissues | EPF associated; also PV associated, with a contrary OR |
| *rs3129949*A* | C>A | 21% EUR; 31% AFR; 15% AMR | Intron 4 TSBP1-AS1; intron 18 TSBP1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region | eQTL for HCG23, HLA-DRB5, HLA-DQB1, C4A, C4B and others in skin tissues; sQTL for HLA-DQA2 and -DPB2 in skin tissues | EPF associated; also PV associated, with a contrary OR |
| *rs9268103*A* | G>A | 15% EUR; 22% AFR; 13% AMR | Intron 4 TSBP1-AS1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region | eQTL for HCG23, HLA-DQB1, -DRB5, -DRA, C4A, C4B and others in skin tissues; sQTL for HLA-DRB1, -DRB5, -DRB6, -DQB1 and -DQB2 in skin tissues | EPF associated; also PV suggestive of association, with a contrary OR |
| *rs1634703*A* | G>A | 1% EUR; 17% AFR; 6% AMR | Intron 1 HCG27:6 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region | eQTL for HCG27 and AL645933.3 in skin tissues; sQTL for MICA and NFKBIL1 in skin tissues | EPF associated |
| *rs79792575*A* | C>T | 3% EUR; 6%AFR; 7%AMR | Intron 2 HCG21; intron 4 SFTA2; exon 5 MUCL3 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region | eQTL in non-skin tissues; sQTL for NFKBIL1 in skin tissues | EPF associated |
| *rs3132931*C* | T>G | 22% EUR; 22% AFR; 10% AMR | Intron 4 TSBP1-AS1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region | Regulome score: 1f. TF binding sites for ELK1::HOXA3; MEIS1::DRGX; promoter flanking region; eQTL for HCG23, TSBP1, HLA-DRB1, -DRB5, -DRB6, -DQB1, -C, C4A, C4B and others in skin tissues; sQTL for HLA-DQB1, -DQB2, -DRB6, -DPB1 and -DPB2 in skin tissues | EPF associated |

| | | | | | | |
|---|---|---|---|---|---|---|
| *rs9468829*C* | T>G | 15% EUR; 27% AFR; 14% AMR | Intron 2 HCG20 | LD (r2 > 0.8) with HCG20 SNPs | eQTL for HCG27 and GTF2H4 in skin tissues; sQTL in non-skin tissues | EPF associated |
| *rs1003879*A* | G>A | 35% EUR; 53% AFR; 35% AMR | Intron 4 TSBP1-AS1; intron 18 TSBP1 | No SNPs in LD (r2 < 0.8) | Binding sites for many TFs; promoter flanking region; eQTL for HLA-DQB1, -DRA, -DRB5, C4A, HCP5 and others in skin tissue; sQTL for HLA-DQB1, -DQB2 and -DQA2 in skin tissues | PV associated |
| *rs3115562*T* | C>T | 6% EUR; 19% AFR; 8% AMR | Intron 4 TSBP1-AS1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region and HLA-DRA | eQTL and sQTL in non-skin tissues | PV associated |
| *rs9268267*G* | A>G | 6% EUR; 20% AFR; 8% AMR | Intron 4 TSBP1-AS1; intron 11 TSBP1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region and HLA-DRA | eQTL and sQTL in non-skin tissues | PV associated |
| *rs498422*G* | T>G | 6% EUR; 32% AFR; 10% AMR | Intron 4 TSBP1-AS1; intron 24 TSBP1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region and HLA-DRA | eQTL for MIR6891 in skin tissues; sQTL in non-skin tissues | PV associated |
| *rs6910071*G* | A>G | 19% EUR; 4% AFR; 23% AMR | Intron 4 TSBP1-AS1; intron 26 TSBP1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region | Regulome score: 1f; eQTL for HLA-DRB1, -DRB6, -DQA1, -DQA2, -DQA2, C4A and others in skin tissue; sQTL for HLA-B, HLA-DQA1 and -DQA2 in skin tissues | PV associated |
| *rs570963*G* | A>G | 14% EUR; 42% AFR; 13% AMR | Intron 4 TSBP1-AS1; intron 24 TSBP1 | LD (r2 > 0.8) with 2 SNPs in TSBP1-AS1/ TSBP1 gene region | eQTL for HLA-DRB5 in skin tissues; sQTL in non-skin tissues | PV associated |
| *rs2273017*G* | T>G | 43% EUR; 74% AFR; 41% AMR | Intron 6 TSBP1-AS1; intron 2 TSBP1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region | Binding sites for many TFs; regulatory region variant; eQTL for HLA-DRB1, -DRB5, -DRB6, -DQB1, -DQA2, C4A and others in skin tissue; sQTL for HLA-DRB1, -DRB5, -DRB6, -DQB1, -DQB2 and -DQA2 in skin tissues | PV associated |
| *rs3129949*A* | C>A | 21% EUR; 31% AFR; 15% AMR | Intron 4 TSBP1-AS1; intron 18 TSBP1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region | eQTL for HCG23, HLA-DQB1, -DRB5, C4A, C4B and others in skin tissues; sQTL for HLA-DQA2 and -DPB2 in skin tissues | PV associated; also EPF associated, with a contrary OR |
| *rs3117137*T* | C>T | 21% EUR; 32% AFR; 15% AMR | Intron 4 TSBP1-AS1; intron 12 TSBP1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region | eQTL for HLA-DQB1, -DRB5, HCP5, C4A, C4B and others in skin tissues; sQTL for HLA-DQA2 in skin tissues | PV associated |

| | | | | | | |
|---|---|---|---|---|---|---|
| *rs521828*T* | <u>C</u>>T | 29% EUR; 55% AFR; 19% AMR | Intron 4 TSBP1-AS1; intron 20 TSBP1 | LD (r2 > 0.8) with SNPs in TSBP1-AS1/ TSBP1 gene region | Splice acceptor variant; eQTL for HLA-DQB1, -DRB5, MICB, C4A, C4B and others in skin tissues; sQTL for HLA-DQA1, -DPB1 and -DPB2 in skin tissues | PV associated |

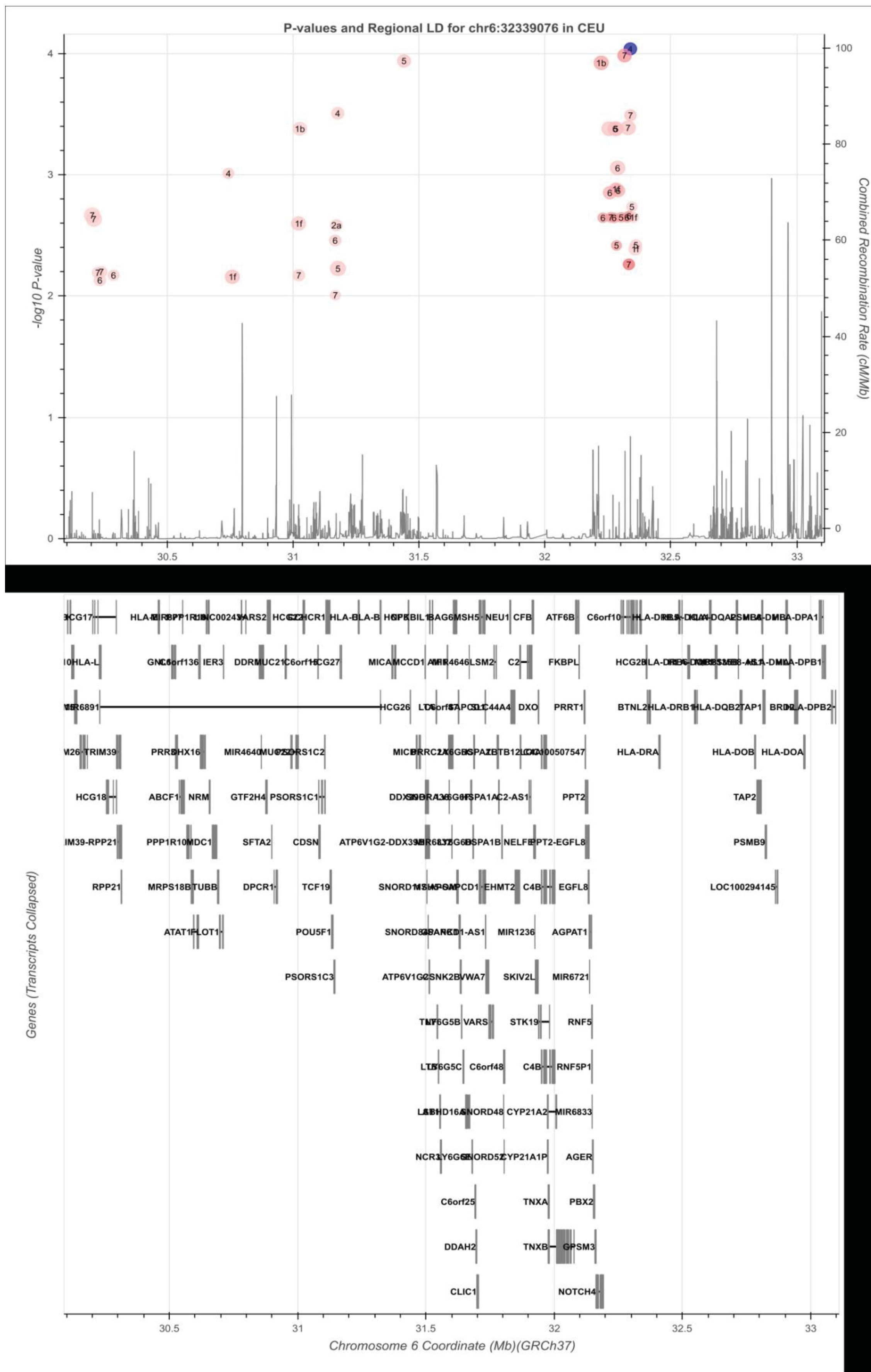
Functional annotations of the 9 PF-associated SNPs and of the PV 10 most-associated SNPs. The underlined alleles are the ancestral ones. MAF - minor allele frequency; EUR - European populations; AFR - African populations; AMR - Amerindian populations; LD - linkage disequilibrium; TF - transcription factor; eQTL - expression quantitative loci trait; sQTL - splice quantitative loci trait; OR - odds ratio. Sources: Ensembl Genome Browser (hg38), LDLink and GTEX Portal (all websites accessed on February, 2020).

**Supp. Table 4.** Interaction annotations for the most associated lncRNAs.

<table>
<tr><th rowspan="2">Gene</th><th rowspan="2">lncRNA pathway</th><th rowspan="2">RNA targets</th><th rowspan="2">RBP targets</th><th rowspan="2">ceRNA</th><th colspan="5">miRNA targets</th></tr>
<tr><th>miRNA</th><th>Target Site</th><th>Class</th><th>Alignment</th><th>miRNA:mRNA targets Pathway</th></tr>
<tr><td rowspan="3">TSBP1-AS1</td><td rowspan="3">-</td><td rowspan="3">THOC5, NIPSNAP1, AC012442.2, FRK, AL591030.1, C10orf143, and others.</td><td rowspan="3">TAF15, WTAP, UPF1, ELAVL1, IGF2BP2, FUS, NOP58, METTL14, ADAR, FBL, EWSR1</td><td rowspan="3">(1) TSBP1-AS1 : miR-552-3p/miR-3064-5p : LINC00514 (LINC00514 pathway: Lysosome)<br><br>(2) TSBP1-AS1 : miR-552-3p/miR-3064-5p : AC005301.1</td><td>hsa-miR-552-3p</td><td>chr6:32230509-32230533[+]</td><td>7mer-m8</td><td>Target: 5' gcuUCGACCCAGUGGGGCACCUGUc 3'<br>|| | |||||    |||||||<br>miRNA : 3' aacAGAUUGGUCA----GUGGACAa 5'</td><td>Many, including: adherens junction, focal adhesion, B cell antigen receptor, endocytosis, and others (includind different signaling pathways)</td></tr>
<tr><td>hsa-miR-3064-5p</td><td>chr6:32230523-32230547[+]</td><td>8mer</td><td>Target: 5' ggGCACCUGUCAGAUGCCAGCCAGa 3'<br>||||  ::|  |:: |||||||<br>miRNA : 3' aaCGUG--UGG--UGUUGUCGGUCu 5'</td><td>Many, including: Signaling by Tgfb receptor complex, immune system, adaptative immune system, and others (includind different signaling pathways)</td></tr>
<tr><td>hsa-miR-6504-5p</td><td>chr6:32230527-32230547[+]</td><td>8mer</td><td>Target: 5' accUGUCAGAUGC-CAGCCAGa 3'<br>::| | | || |||||||<br>miRNA : 3' gacGUAAUGU-CGUGUCGGUCu 5'</td><td>Many, including: apoptosis, T cell receptor signaling, focal adhesion, and others (includind different signaling pathways)</td></tr>
<tr><td>HCG27</td><td>Natural killer cell mediated cytotoxicity, p38 signaling pathway, others</td><td>MICB, MICA, HLA-F-AS1, MICE, C1QBP, EIF2D, ZFP64, CCM2, MICC</td><td>U2AF2, TAF15, PRPF8, FAM120A, SRSF1, SLTM, HNRNPU, HNRNPK, CSTF2T, HNRNPA1, CSTF2T, SLTM, HNRNPUL1, SND1, ELAVL1, FXR2, FXR1, FUS, and others.</td><td>-</td><td colspan="5">-</td></tr>
</table>

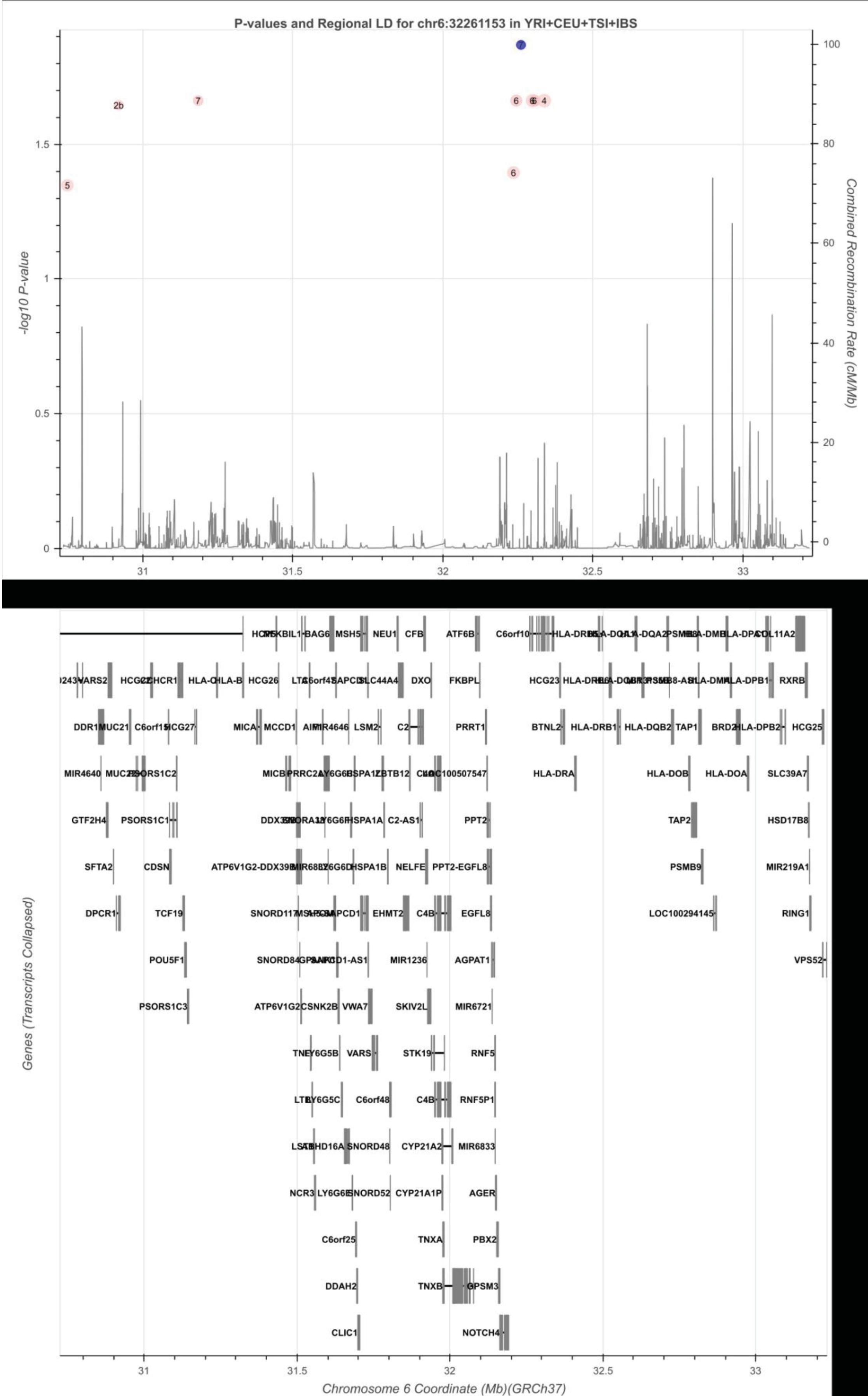
Interaction and pathway annotations for the TSBP1-AS1 and HCG27 lncRNA genes. RBP - RNA binding protein. CeRNA - competing endogenous RNA. Sources: ENCORI Browser (available in http://starbase.sysu.edu.cn/index.php)

**Suppl. Figure 1.** LD region of the most associated SNPs in PV.




The most PV associated SNP is represented by a purple dot, while the other associated SNPs are represented by pink dots. Among the strongly linked SNPs (r2 > 0.8, represented on the right side as "combined recombination rates" higher than 80%), all of them are included in the associated gene region (genes represented below, according to their genomic locations in chromosome 6). Source: LDLink (accessed on February, 2020).

**Suppl. Figure 2.** LD region of the EPF-associated SNPs.






The most EPF associated SNP is represented by a purple dot, while the other associated SNPs are represented by pink dots. Among the strongly linked SNPs (r2 > 0.8, represented on the right side as "combined recombination rates" higher than 80%), all of them are included in the associated gene region (genes represented below, according to their genomic locations in chromosome 6). Source: LDLink (accessed on February, 2020).

## 6. CAPÍTULO II (Manuscrito de qualificação)

ORIGINAL ARTICLE

**GENETIC VARIABILITY OF IMMUNE-RELATED LNCRNAS: THE EFFECT OF POLYMORPHISMS IN *LINC-PINT* AND *LY86-AS1* ON PEMPHIGUS FOLIACEUS SUSCEPTIBILITY**

***Running title:*** Immune-related lncRNAs in pemphigus foliaceus

Amanda Salviano-Silva[1], Ticiana Della Justina Farias[1], Valéria Bumiller-Bini[1], Mariana de Sousa Castro[1], Sara Cristina Lobo-Alves[1], Hauke Busch[2,#], Margitta Worm[3], Matthias Goebeler[4], Nina van Beek[5], Andre Franke[6], Michael Wittig[6], Danillo Gardenal Augusto[1,##], Detlef Zillikens[5], Rodrigo Coutinho de Almeida[1,###], Angelica Beate Winter Boldt[1], Maria Luiza Petzl-Erler[1], Enno Schmidt[2,5], Danielle Malheiros[1].

[1] *Postgraduate Program in Genetics. Department of Genetics, Federal University of Paraná (UFPR), Curitiba, Brazil.*

[2] *Lübeck Institute of Experimental Dermatology, University of Lübeck, Lübeck, Germany.*

[3] *Division of Allergy and Immunology, Department of Dermatology, Venerology and Allergy, Charité-Universitätsmedizin Berlin, Berlin, Germany.*

[4] *Department of Dermatology, Venereology and Allergology, University Hospital Würzburg, Würzburg, Germany.*

[5] *Department of Dermatology, University of Lübeck, Lübeck, Germany.*

[6] *Institute of Clinical Molecular Biology (IKMB), Christian-Albrechts-University of Kiel, Kiel, Germany.*

Current address:

[#] *Institute for Cardiogenetics, University of Lübeck, Lübeck, Germany.*

[##] *Department of Neurology, University of California San Francisco, USA.*

[###] *Department of Biomedical Data Sciences, Sector Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands.*

**Corresponding author**

Prof. Dr. Danielle Malheiros

**Postal address:**

Laboratory of Human Molecular Genetics, Department of Genetics, Federal University of Paraná, Av. Cel. Francisco H. dos Santos, 100, Zip Code 81530-000, Curitiba, PR, Brasil

**Telephone:** +55 41 3361 1724

E-mail: dani_malheiros@ufpr.br

**Funding sources:** Fundação Araucária (FA), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Finance code 001), Doctorate Sandwich Exchange Program (PDSE/CAPES). **The authors have no conflict of interest to declare.**

Main text word count (from introduction to conclusion): 3,359

Full text word count: 6,858

Number of references: 56

Number of tables: 3 (not including supplemental)

Number of figures: 2 (not including supplemental)

**Abbreviations:** *CASC15*, cancer susceptibility 15; *cDNA*, complementary DNA; *CEU*, Utah Residents with Northern and Western European Ancestry; *Chr*, chromosome; *CI 95%*, 95% confidence interval; *Cq*, quantification cycle; *DSG1*, desmoglein-1; *eQTL*, expression quantitative trait loci; *H19*, imprinted maternally expressed transcript; *IBS*, Iberian Population in Spain; *IFNG-AS1*, interferon-gamma - antisense RNA 1; *IL16*, interleukin 16; *LINC-PINT*, lincRNA - p53 induced transcript; *LD*, linkage disequilibrium; *lnc-C5AR1*, lncRNA - complement component 5a receptor 1; *lnc-IL17B-5*, lncRNA - interleukin 17B transcript 5; *lnc-LINGO2-5*, lncRNA - leucine rich repeat and Ig domain containing 2 - transcript 5; *lncRNA*, long non-coding RNA; *LY86-AS1*, lymphocyte antigen 86 - antisense RNA 1; *MAF*, minor allele frequency; *MEG3*, maternally expressed 3; *NEAT1*, nuclear paraspeckle assembly transcript 1; *OR*, odds ratio; *PBMC*, peripheral blood mononuclear cells; *PF*, pemphigus foliaceus; *PC*, principal component; *PCA*, principal component analysis; *qPCR*, quantitative polymerase chain reaction; *SLE*, systemic lupus erythematosus; *SNP*, single nucleotide polymorphism; *SRE*, splicing regulatory elements; *TCGA*, the cancer genome atlas; *TSI*, Toscani in Italia.

## ABSTRACT

**Background:** Pemphigus foliaceus (PF) is an autoimmune blistering disease of the skin, clinically characterized by erosions and, histopathologically, by acantholysis. PF is endemic in the Brazilian Central-Western region. Numerous single nucleotide polymorphisms (SNPs) have been shown to affect the susceptibility for PF, including SNPs at long non-coding RNA (lncRNAs) genes, which are known to participate in many physiological and pathogenic processes, such as autoimmunity. **Objective:** We investigated whether the genetic variation of immune-related lncRNA genes affects the risk for endemic and sporadic forms of PF. **Methods:** We analyzed 692 novel SNPs for PF from 135 immune-related lncRNAs genes in 227 endemic PF patients and 194 controls. The SNPs were genotyped by Illumina microarray and analyzed by applying logistic regression at additive model, with correction for sex and population structure. Four associated SNPs were also evaluated in an independent German cohort of 76 sporadic PF patients and 150 controls. Further, we measured the LINC-PINT and LY86-AS1 levels by quantitative PCR in peripheral blood mononuclear cells of healthy subjects. **Results:** We found 27 SNPs associated with endemic PF ($p<0.05$ without overlapping with protein-coding genes), of which we highlight *LY86-AS1 rs12192707*A* (OR=0.68, $p$=0.014) and *LINC-PINT rs10228040*A* (OR=1.47, $p$=0.012). The SNP *rs10228040*A* was also associated with increased susceptibility for sporadic PF (OR=2.28, $p$=0.002). Moreover, the *A+* carriers of *LY86-AS1*rs12192707* mark lowest LY86-AS1 levels, which may be associated with a decreasing autoimmune response. **Conclusion:** our results suggest a critical role of lncRNA variants in immunopathogenesis of both PF endemic and sporadic forms.



**Key-words:** lncRNAs, pemphigus foliaceus, association study, LINC-PINT, LY86-AS1.

## 1. INTRODUCTION

Pemphigus foliaceus (PF) is a severe autoimmune blistering disease characterized by the presence of autoantibodies that recognize desmosomal antigens, especially desmoglein-1 (DSG1), at the surface of keratinocytes [1,2]. The pathogenic auto-recognition of these adhesion molecules is related to keratinocyte detachment, a process known as acantholysis, whose clinical manifestation is epithelial blistering [1–7]. PF is sporadic and rare worldwide*, including in Europe and developed countries. However, it is frequently observed in countries where PF is considered endemic, such as Brazil, Peru and Tunisia.* In Brazil, PF endemicity occurs in the central-western region, where the disease is popularly known as *fogo selvagem* (which means "wild fire" in Portuguese), reaching its worldwide highest incidence (approximately 3.4% in a Terena indigenous community - *Limão-Verde* reservation, MS/Brazil) [8–11]. PF is a multifactorial disorder, whose etiology is associated with genetic and epigenetic factors, interacting with specific environmental conditions and modulating the susceptibility of the individual to the disease. The activation of autoreactive T and B cells culminates with the production of specific IgG autoantibodies against DSG1 and acantholysis [1–3,6]. Interestingly, patients of either endemic or sporadic PF present similar clinical, histological and immunological aspects. Individuals living in endemic areas commonly exhibit IgM antibodies against DSG1, which suggests an immunological response against environmental antigens during the preclinical stage [12].

Variants within numerous potential susceptibility genes have been analyzed in the context of PF, mainly in case-control studies. The majority of the genes associated with differential susceptibility to PF are responsible for coding proteins involved in immune responses, thus being related to the autoimmune and autoinflammatory features of pemphigus [13,14,23–26,15–22]. Besides, there is growing evidence that variants located in non-coding regions, such as in genes of long non-coding RNAs (lncRNAs), also influence the susceptibility to PF [26,27].

LncRNAs are non-coding transcripts with more than 200 nucleotides that are strictly regulated and mainly involved in distinct transcriptional and post-transcriptional regulatory processes [28–30]. A growing body of evidence indicates that genetic variations may alter the structure or expression levels of lncRNAs thus resulting in modulation of many physiological [30] and pathological [31,32] conditions. As a consequence, it is not surprising that lncRNAs have

been recognized to be involved in several human diseases, including autoimmune disorders [31,33].

Aiming to explore the involvement of immune-related lncRNAs in PF susceptibility, we investigated 692 single nucleotide polymorphisms (SNPs) located in lncRNA genes selected due to their involvement in immune responses. We expanded our results by interrogating if certain variants associated with endemic PF are also associated in an independent German cohort of sporadic PF. Our results confirm previous findings from our group [27] that point to a major role of lncRNAs modulating PF susceptibility. We found that the SNP *rs10228040* (*LINC-PINT* gene) is associated with susceptibility for both endemic and sporadic forms of PF, while *rs12192707* (*LY86-AS1* gene) is associated with protection against endemic PF and marks *LY86-AS1* expression levels. Furthermore, the associated SNPs are predicted to influence alternative gene splicing, and co-expression annotations of LINC-PINT and LY86-AS1 include potential immunological roles of these lncRNAs in pemphigus autoimmunity.

## 2. MATERIAL AND METHODS

### 2.1 Study populations and samples

All subjects voluntarily agreed to participate in this study and signed an informed consent form, in accordance with the Declaration of Helsinki. This study was performed according to Brazilian and German federal laws and was approved both by the National Committee for Ethics in Research (CONEP protocol CAAE 02727412.4.0000.0096, approval 505.988) and the Ethics Committee of the University of Lübeck (08-156, 12-178). The study was performed in cooperation with the Biobank PopGen, Kiel, Germany.

Patients were diagnosed based on physical examination combined with immunological testing, histopathology and/or immunohistochemistry of skin biopsies. All individuals were unrelated, and those with known history of any other autoimmune disease were excluded from the study.

#### 2.1.1 Endemic PF cohort

The endemic PF cohort comprised 227 PF patients (52% female) and 194 controls (51.3% female) from endemic areas in Brazil. All individuals were of predominantly European ancestry, and were contacted at reference hospitals located in the endemic areas, which were previously reported in [34]. DNA was extracted from peripheral blood of endemic PF cohort subjects by the phenol-chloroform-isoamyl alcohol method [35] and stored at -80°C.

#### 2.1.2 Sporadic PF cohort

The sporadic PF cohort was collected in Germany hospitals by the German Autoimmune Bullous Diseases Genetic Study Group and comprised 76 PF patients (46% female) and 150 controls (50% female), all Caucasoid Europeans. Patients were diagnosed based on a compatible clinical phenotype and a positive direct immunofluorescence microscopy of a perilesional skin biopsy or serum autoantibodies against DSG1 as detected by ELISA (Euroimmun, Lübeck, Germany) according to the guideline of the German Society of Dermatology [36,37]. EDTA blood was stored at -80°C until processed. DNA was extracted from peripheral blood using the QIAamp DNA Maxi Blood Kit (Qiagen, Hilden, Germany) and stored at -80°C until genotyping.

### 2.1.3 Expression analysis cohort

For quantification of lncRNA levels, we analyzed a panel of healthy individuals from the city of Curitiba (Paraná state, Brazil) and its metropolitan region. All individuals were of predominantly European ancestry. We excluded those who (i) were pregnant; (ii) reported chronic or recent acute health alterations, including allergies, and those who were under the use of prescribed or over-the-counter medications; (iii) consumed alcohol, tobacco and other drugs less than 48h before the blood collection.

RNA samples were isolated from subjects' peripheral blood mononuclear cells (PBMCs) by SV Total RNA Isolation System kit (Promega Corporation, USA). cDNA synthesis were performed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems®, USA).

## 2.2 Microarray genotyping

For the endemic PF cohort, genotyping was performed with microarray DNA Human Infinium® CoreExome-24 Beadchip (Illumina®, San Diego, USA), according to the manufacturer's instructions.

## 2.3 Selection of lncRNA candidates

We selected 135 lncRNAs (Table S1) for genetic variability investigation in endemic PF. The selection was based on their potential immunological roles, according to the following criteria: (i) lncRNAs cited in the scientific literature (research articles available in PubMed/NCBI until 12/2018) to be involved in homeostasis of immune cells and/or autoimmune diseases (articles searching according to key-words related to lncRNAs, immune cells homeostasis, autoimmune diseases); and/or (ii) lncRNAs that were antisense to, neighboring and/or related to protein-coding genes previously associated with pemphigus (foliaceus and/or vulgaris) and/or immunological processes, according to literature and databases. None of these lncRNAs were previously investigated in any subtype of pemphigus.

## 2.4 Genotyping data analysis

Genomic regions of the selected immune-related lncRNA genes were searched at LNCiPedia DB (GRCh37/hg19) [38] to extract 1,113 tag autosomal variants from the microarray

genotyping data that mapped to these genes. Manipulation of the SNP data was performed with PLINK version 1.09 [39]. For the quality control, markers with a genotyping call rate < 96% were excluded, and principal component analysis (PCA) was used to control for possible population structure. We excluded variants having (i) a minor allele frequency (MAF) < 1%, (ii) a genotype distribution deviating from the Hardy-Weinberg equilibrium in the control sample ($p < 0.05$) and (iii) a strong linkage disequilibrium ($r^2 > 0.8$) with other SNPs in this analysis. A total of 692 SNPs of immune-related lncRNAs remained for further analysis (Table S2).

We performed an association analysis at additive, dominant and recessive models by logistic regression with odds ratio (OR) and confidence interval of 95% (CI 95%), applying corrections for sex and two principal components (PCs). The limit of significance adopted was $p \leq 0.05$.

2.5 In silico analysis

We selected the four SNPs with strongest association (p < 0.015) for *in silico* analysis. The four SNPs were carefully evaluated for their linkage disequilibrium (LD) with other SNPs within the European sample (CEU - Utah Residents with Northern and Western European Ancestry; TSI - Toscani in Italia; IBS - Iberian Population in Spain) from the 1000 Genomes project data, using LDLink [40] online tool. Functional annotations and the predicted regulatory impact, such as transcription factors binding and alternative gene splicing, were analyzed using Ensembl [41] and Human Splicing Finder [42] browsers.

Furthermore, we explored functional annotation for the four lncRNAs, in which the strongest associations were found. We searched for annotated information for co-expressed genes, gene ontology and enriched pathways using the circlncRNAnet database demo server [43], which comprises co-regulatory annotations for different tumor tissues from The Cancer Genome Atlas (TCGA) data [44].

2.6 Evaluation of lncRNA genotyping in sporadic PF

In order to check whether the associations identified in endemic PF are representative for PF in general or dependent from epidemiological factors solely relevant for endemic PF, we selected 4 associated SNPs for evaluation in a sporadic PF cohort by a second genotyping

method, using the MassARRAY iPlex platform (Agena Biosciences, USA). The primer and extension sequences for the evaluated SNPs are listed in Table S6. Moreover, these SNPs were also genotyped in 195 patients and 138 controls from the endemic cohort, as an internal genotyping quality control to evaluate the genotyping concordance with the first method.

The SNPs were analyzed by logistic regression applying correction for sex, with additive, dominant and recessive models, using Plink Program v.1.09 [39]. The limit of significance adopted was $p \leq 0.05$.

2.7 Quantification of lncRNA relative expression levels

Quantitative real-time polymerase chain reaction (qPCR) was used to measure the expression levels of selected lncRNAs in PBMCs of healthy subjects. To this end, we selected the genotypes for *LINC-PINT rs10228040*A>G* (*A/A*=11, *A/G*=11, *G/G*=16) and *LY86-AS1 rs12192707*A>G* (*A/A*=6, *A/G*=8, *G/G*=10). The qPCR reactions were performed in triplicates using GoTaq® qPCR Master Mix (Promega Corporation, USA) with specific primers on exon-exon junctions (Table 1), using the ViiA 7 Real Time PCR System (Applied Biosystems®, USA). All primer pairs presented similar efficiency and the amplifications presented unique peaks in the melting curves (data not shown).

The gene expression levels were normalized by the expression levels of the housekeeping gene *GAPDH*. Fold changes were calculated by the comparative quantification cycle (Cq) method $2^{-\Delta\Delta Cq}$ [45]. We performed the Mann-Whitney test for Cq differences, using GraphPad Prism v.6 software. The limit of significance adopted was $p \leq 0.05$.

## 3. RESULTS

3.1 SNPs in immune-related lncRNA genes are associated with endemic PF

We found 47 SNPs located at lncRNA genes related to immune responses associated with endemic PF ($p < 0.05$ (Table S3)). Out of these, we focused on the 27 associated SNPs located only in lncRNA genes and not overlapping with protein-coding genes, in order to avoid the bias of coding-gene associations (Table 2).

These 27 SNPs uniquely mapping to lncRNA were located in *CASC15* (6 associated SNPs), *LINC-PINT* (5 associated SNPs), *LY86-AS1* (3 associated SNPs), *lnc-LINGO2-5* (6 associated SNPs), *LINC01991* (1 associated SNP), *NEAT1* (1 associated SNP), *MEG3* (1 associated SNP), *IFNG-AS1* (1 associated SNP), *lnc-C5AR1* (1 associated SNP), *H19* (1 associated SNP) and *lnc-IL17B-5* (1 associated SNP) genes [38]. All of them occur in lncRNA intronic regions, except *rs674485*, which is located at the single *NEAT1* exon. The top four strongest signals were: *CASC15*rs2473122* ($p$ = 0.01, OR = 1.721), *LINC-PINT*rs10228040* ($p$ = 0.012, OR = 1.465), *LINC-PINT*rs7812207* ($p$ = 0.013, OR = 0.614), and *LY86-AS1*rs12192707* ($p$ = 0.014, OR = 0.68) (Table 2).

These results suggest that the genetic variability of immune-related lncRNAs genes influences the susceptibility to PF.

3.2 Computational and in silico results

We searched for genetic annotations and applied *in silico* analysis for the top-4 associated SNPs (*CASC15*rs2473122*, *LINC-PINT*rs10228040* and **rs7812207*, and *LY86-AS1*rs12192707*) in order to predict their possible impact on lncRNA regulation or interactions with other biological molecules (Table S4).

The SNPs rs2473122, rs10228040 and rs12192707 are in strong linkage disequilibrium (LD) ($r^2 > 0.8$) with other polymorphisms of their respective genes (1000 genomes project data [40], which may be held responsible for the observed association effect. For rs7812207, there are no SNPs presenting strong LD in the analyzed population [40], suggesting that its association with PF results from a possible direct causal effect (Table S4). Moreover, the MAFs of rs10228040 and rs12192707 show a large discrepancy between European and African populations (Table S4) [41],

which, considering the composition of the Brazilian population, might affect the interpretation of genetic associations between sporadic and endemic forms.

The intronic location of the selected SNPs leads us to suggest that they are not likely to affect the lncRNA secondary structure. However, they might be located at splicing sites and modulate the alternative splicing, generating different lncRNA isoforms. To evaluate this possibility, we analyzed allele flanking sequences (+/-100nt) using the Human Splicing Finder [42], an online tool for prediction of variant effects on splicing signals and motifs. The SNPs rs2473122, rs7812207 and rs12192707 were predicted to affect motifs of splicing-regulatory elements (SREs) (Table S4) by creating or blocking splicing sites.

Moreover, using the online tool circlncRNAnet [43], we observed that the lncRNAs genes *CASC15*, *LINC-PINT*, and *LY86-AS1* are co-expressed in many tissues with genes involved in immune response and epithelial function (Table S5). Among these mRNA/lncRNA co-expressed genes, we highlight in normal tissues: *LY86-AS1* is co-expressed with *IL16* (*interleukin 16*), *CD19* and further genes involved in lymphocytic activity; *LINC-PINT* is co-expressed with genes involved in DNA repair and secretory activity; and *CASC15* is co-expressed with collagen genes. Enriched biological processes for *LY86-AS1* include B cell receptor signaling pathways, regulation of chemotaxis and positive regulation of immune and stimulus responses, while *CASC15* is involved in the regulation of B-cell mediated immunity, somatic regulation/diversification of immunoglobulins and DNA recombination (Figure S1).

All of these *in silico* and computational results point to many regulatory mechanisms which might be affected by lncRNA genetic variability, possibly reflecting the involvement of different genes and mechanisms related to PF immunopathogenesis.

### 3.3 LINC-PINT*rs10228040 is associated with sporadic PF

For evaluation in a sporadic PF cohort by a second genotyping method (MassARRAY iPlex platform), we selected the following SNPs: *LINC-PINT*rs10228040* and *LY86-AS1*rs12192707*, which have the highest endemic PF associations and present higher discrepancy between European and African populations, whose admixture is present in endemic regions; and *NEAT1*rs674485* and *MEG3*rs1884537*, whose genes are well known in literature to be associated with many multifactorial diseases [31]. The reproducibility rate was higher than 98.5%

to the first genotyping method (data not shown). Genotype distributions agreed with Hardy-Weinberg equilibrium predictions for controls, and the allele frequency differences between controls and patients were further analyzed by logistic regression.

*LINC-PINT rs10228040*A* was associated with susceptibility to sporadic PF with additive (OR = 2.284, $p$ = 0.0021), dominant (OR = 2.499, $p$ = 0.0055) or recessive (OR = 3.319, $p$ = 0.0484) effect (Table 3). No significant associations were found for rs12192707, rs674485 and rs1884537 in this sporadic cohort.

3.4 LY86-AS1*rs12192707 marks differential expression

To evaluate the impact of the associated SNPs on gene expression, *LINC-PINT* and *LY86-AS1* levels were quantified in PBMCs of non-endemic healthy subjects through qPCR, and the fold changes were analyzed based on genotypes of rs10228040 and rs12192707, respectively.

LY86-AS1 levels were markedly reduced in *rs12192707*A* carriers (*A/A* and *A/G* genotypes), when compared with individuals presenting the *G/G* genotype (Figure 1A). No significant differences between rs10228040 genotypes were seen for the expression of *LINC-PINT* (Figure 1B).

Therefore, *rs12192707*A* is associated with *LY86-AS1* differential expression in PBMCs.

## 4. DISCUSSION

LncRNAs are regulatory transcripts capable of interacting with several biological molecules, thus participating in important physiological and pathological mechanisms. These interactions can be influenced by genetic variability in lncRNA genes, possibly resulting in deregulation of gene networks. A recent study of our group showed interesting genetic associations of lncRNA variants in endemic PF [27], however, without focusing on immune-related lncRNA genes. Here, SNPs located in lncRNA genes previously reported in literature to be involved in immunological processes and diseases were investigated in endemic PF for the first time. Furthermore, associated SNPs found in the present study were also evaluated in a sporadic cohort and investigated with respect to their possible influence on gene expression.

We found several SNPs in immune-related lncRNAs to be associated with endemic PF, including some that overlapped with protein-coding genes and others specific of lncRNA genes. Among the lncRNA-specific SNPs with the lowest *p* values, *CASC15 rs2473122*A* and *LINC-PINT rs10228040*A* were associated with increased susceptibility to endemic PF, while *LINC-PINT rs7812207*A* and *LY86-AS1 rs12192707*A* were associated with protection against the disease. Moreover, the SNPs rs2473122, rs7812207 and rs12192707 are predicted to be located in splicing regulatory elements (SREs) motifs. In this scenario, the *LY86-AS1* SNP *rs12192707*A* was predicted to create exonic splicing enhancers (ESE) sites and exon identity elements (EIE), while to disrupt exonic splicing silencers (ESS) sites, thus suggesting an impact on alternative splicing of the lncRNA transcript.

The main populations composing the Brazilian endemic cohort (CEU, IBS and TSI composing European ancestries and YRI composing African ancestries) have high frequency differences between both ancestries for *LY86-AS1 rs12192707*A,* and especially for *LINC-PINT rs10228040*A*. Therefore, we reevaluated these SNPs in a sporadic PF cohort from Germany to investigate whether the associations are representative for PF in general or depend on epidemiological factors from endemic areas. In addition to these two SNPs, we also included in this second genotyping the *NEAT1*rs674485* and *MEG3*rs1884537*, due to their already known associations with many multifactorial diseases documented in the literature [31,32]. Here, we found that rs10228040*A* (*LINC-PINT*) is also associated with increased susceptibility for sporadic

PF, in contrast to *NEAT1*rs674485*, *MEG3*rs1884537* and *LY86-AS1*rs12192707*, which were only associated with the endemic form of disease.

Interestingly, we observed in PBMCs of healthy subjects that the RNA expression levels of LY86-AS1 are lower for the *rs12192707 A+* genotypes (*A/A* plus *A/G*) than for the *G/G* genotype, suggesting an expression quantitative trait loci (eQTL) effect of this SNP on its host gene. Therefore, *rs12192707*A* seems to be associated with protection against endemic PF by downregulating *LY86-AS1*. Conversely, the expression levels of LINC-PINT were independent of the rs10228040 genotypes in the analyzed samples.

*LY86-AS1* is co-expressed in several tissues with essential genes involved in the B-cell receptor signaling pathway and positive regulation of the immune response, such as *IL16*, *CD19* and *BLK* [43]. Regarding to the gene ontology analysis, the LY86-AS1 correlated genes are also associated with positive regulation of response to stimuli, which could explain the different susceptibilities found between sporadic and endemic forms of PF, regarding the individual immunological response to the (still unknown) environmental factors present in endemic and non-endemic areas. Among the *LY86-AS1* co-expressed genes, variants of *CD19* and *IL16* were previously associated with bullous pemphigoid (another autoimmune bullous skin disease) [46,47] and other autoimmune diseases that affect the skin, such as psoriasis [48,49] and systemic lupus erythematosus (SLE) [50–52]. Thus, it is possible that immunological events that are associated with PF autoimmunity (mainly when induced by environmental factors present in endemic areas) are suppressed by decreased LY86-AS1 levels, which in turn are associated with the protective allele *rs12192707*A*.

Although the rs10228040 genotypes show no differential expression of *LINC-PINT* in PBMC samples, we found SNP associations in different models with both endemic and sporadic PF forms, as well as its LD SNPs located in the same lncRNA gene. This suggests that LINC-PINT may be an important genetic factor in PF etiology. *LINC-PINT* is co-expressed in many tissues with various genes, including *SCAMP2* and *ARHGAP17* [43]. SCAMP2 is involved in vesicular transport as well as cytokine and chemokine secretion and is highly expressed in mast cells [53]. Furthermore, SCAMP2 is a candidate gene within associated *loci* for SLE [54], and its expression is altered in response to glucocorticoids [55]. ARHGAP17, a RhoGTPase-activating protein also known as Rich1, plays a role in cell adhesion by regulating the formation and maintenance of

tight epithelial junctions and adherence junctions [56]. Thus, LINC-PINT might be involved in gene regulatory networks potentially related to inflammatory mechanisms and adhesion loss, which are important for the pathogenesis of PF.

Taken together, we conclude that the genetic variability of immune-related lncRNAs influences the susceptibility to PF, which might differ between the endemic and sporadic forms of the disease. We highlight the lncRNAs *LY86-AS1* and *LINC-PINT*, which might be involved in the pathogenesis of PF and other immune-related skin diseases. The SNPs and lncRNAs identified in this study are suggested as good candidates for further functional investigation, which could help to better understand the molecular mechanisms involved in the immunopathogenesis of PF and, possibly, of other bullous autoimmune diseases.

**ACKNOWLEDGEMENTS**

We gratefully acknowledge the patients for their participation in this study. We also thank the staff of the Laboratório de Genética Molecular Humana/UFPR for their assistance and helpful discussions, and the principal investigators of the German Autoimmune Bullous Diseases Study Group, who contributed with less than five patients to this study and who were not granted co-authorships, i.e. Dr. Claudia Günther (Dresden), Dr. Michael Sticherling (Erlangen), Dr. Eva Hadaschik (Heidelberg), Dr. Kerstin Steinbrink (Mainz), Dr. Regine Gläser (Kiel) and Dr. Wiebke Ludwig-Peitsch (Mannheim).

This work was supported by grants of Fundação Araucária (FA), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Doctorate Sandwich Exchange Program (PDSE/CAPES), Co-financed Short-Term Research Grant Brazil (2018) by Deutscher Akademischer Austauschdienst (DAAD), and the German Research Foundation under Germany’s Excellence Strategy (EXC 2167-390884018; HB, DZ, ES). We report that there are no known conflicts of interest associated with this publication.

## REFERENCES


[1] E. Schmidt, M. Kasperkiewicz, P. Joly, Pemphigus., Lancet (London, England). 394 (2019) 882–894. https://doi.org/10.1016/S0140-6736(19)31778-7.

[2] M. Kasperkiewicz, C.T. Ellebrecht, H. Takahashi, J. Yamagami, D. Zillikens, A.S. Payne, M. Amagai, Pemphigus, Nat. Rev. Dis. Prim. 3 (2017) 17026. https://doi.org/10.1038/nrdp.2017.26.

[3] R. Pollmann, T. Schmidt, R. Eming, M. Hertl, Pemphigus: a Comprehensive Review on Pathogenesis, Clinical Presentation and Novel Therapeutic Approaches, Clin. Rev. Allergy Immunol. 54 (2018). https://doi.org/10.1007/s12016-017-8662-z.

[4] V. Spindler, R. Eming, E. Schmidt, M. Amagai, S. Grando, M.F. Jonkman, A.P. Kowalczyk, E.J. Müller, A.S. Payne, C. Pincelli, A.A. Sinha, E. Sprecher, D. Zillikens, M. Hertl, J. Waschke, Mechanisms Causing Loss of Keratinocyte Cohesion in Pemphigus, J. Invest. Dermatol. 138 (2018) 32–37. https://doi.org/10.1016/j.jid.2017.06.022.

[5] M. Takeichi, Cadherin cell adhesion receptors as a morphogenetic regulator., Science. 251 (1991) 1451–5. https://doi.org/10.1126/science.2006419.

[6] D.A. Culton, Y. Qian, N. Li, D. Rubenstein, V. Aoki, G.H. Filhio, E.A. Rivitti, L.A. Diaz, Advances in pemphigus and its endemic pemphigus foliaceus (Fogo Selvagem) phenotype: A paradigm of human autoimmunity, J. Autoimmun. 31 (2008) 311–324. https://doi.org/10.1016/j.jaut.2008.08.003.

[7] R.M. Nisihara, R.S. de Bem, R. Hausberger, V.S. Roxo, D.P. Pavoni, M.L. Petzl-Erler, I.J. de Messias-Reason, Prevalence of autoantibodies in patients with endemic pemphigus foliaceus (fogo selvagem)., Arch. Dermatol. Res. 295 (2003) 133–7. https://doi.org/10.1007/s00403-003-0412-9.

[8] A.M. Abréu-Vélez, I.J. de M. Reason, M.S. Howard, A.M. Roselino, Endemic pemphigus foliaceus over a century: Part I., N. Am. J. Med. Sci. 2 (2010) 51–9. http://www.ncbi.nlm.nih.gov/pubmed/22624115 (accessed September 17, 2019).

[9] J.-C. Bystryn, J.L. Rudolph, Pemphigus., Lancet (London, England). 366 (n.d.) 61–73. https://doi.org/10.1016/S0140-6736(05)66829-8.

[10] L.A. Diaz, S.A.P. Sampaio, E.A. Rivitti, C.R. Martins, P.R. Cunha, C. Lombardi, F.A. Almeida, R.M. Castro, M.L. Macca, C. Lavrado, G.H. Filho, P. Borges, A. Chaul, L. Minelli, J.C. Empinotti, H. Friedman, I. Campbell, R.S. Labib, G.J. Anhalt, Endemic pemphigus foliaceus (Fogo selvagem): II. Current and historic epidemiologic studies, J. Invest. Dermatol. 92 (1989) 4–12. https://doi.org/10.1111/1523-1747.ep13070394.

[11] G. Hans-Filho, V. Aoki, E. Rivitti, D.P. Eaton, M.S. Lin, L.A. Diaz, Endemic pemphigus foliaceus (fogo selvagem)--1998. The Cooperative Group on Fogo Selvagem Research., Clin. Dermatol. 17 (1999) 225–35; discussion 105-6. https://doi.org/10.1016/S0738-081X(99)00014-0.

[12] L.A. Diaz, P.S. Prisayanh, D.A. Dasher, N. Li, F. Evangelista, V. Aoki, G. Hans-Filho, V. dos Santos, B.F. Qaqish, E.A. Rivitti, Cooperative Group on Fogo Selvagem Research, The IgM anti-desmoglein 1 response distinguishes Brazilian pemphigus foliaceus (fogo selvagem) from other forms of pemphigus., J. Invest. Dermatol. 128 (2008) 667–75. https://doi.org/10.1038/sj.jid.5701121.

[13] M.L. Petzl-Erler, J. Santamaria, Are HLA class II genes controlling susceptibility and resistance to Brazilian pemphigus foliaceus (fogo selvagem)?, Tissue Antigens. 33 (1989) 408–14. https://doi.org/10.1111/j.1399-0039.1989.tb01684.x.

[14] J.D. Torzecka, J. Narbutt, A. Sysa-Jedrzejowska, M. Borowiec, A. Ptasinska, G. Woszczek, M.L. Kowatski, Tumour necrosis factor-α polymorphism as one of the complex inherited factors in pemphigus, Mediators Inflamm. 12 (2003) 303–307. https://doi.org/10.1080/0962935031000161973

[15] N.F. Pereira, J.A. Hansen, M.-T. Lin, V.M.M.S. Roxo, K. Braun, M.L. Petzl-Erler, Cytokine gene polymorphisms in endemic pemphigus foliaceus: a possible role for IL6 variants., Cytokine. 28 (2004) 233–41. https://doi.org/10.1016/j.cyto.2004.08.006.

[16] D.P. Pavoni, L.B. Cerqueira, V.M.M.S. Roxo, M.L. Petzl-Erler, Polymorphism of the promoter region and exon 1 of the CTLA4 gene in endemic pemphigus foliaceus (fogo selvagem)., Brazilian J. Med. Biol. Res. = Rev. Bras. Pesqui. Medicas e Biol. 39 (2006) 1227–32. https://doi.org/10.1590/s0100-879x2006000900010.

[17] K. Braun-Prado, M.L. Petzl-Erler, Programmed cell death 1 gene (PDCD1) polymorphism and pemphigus foliaceus (fogo selvagem) disease susceptibility, Genet. Mol. Biol. 30 (2007) 314–321. https://doi.org/10.1590/S1415-47572007000300003.

[18] D. Malheiros, M.L. Petzl-Erler, Individual and epistatic effects of genetic polymorphisms of B-cell co-stimulatory molecules on susceptibility to pemphigus foliaceus, Genes Immun. 10 (2009) 547–558. https://doi.org/10.1038/gene.2009.36.

[19] R. Dalla-Costa, M.R. Pincerati, M.H. Beltrame, D. Malheiros, M.L. Petzl-Erler, Polymorphisms in the 2q33 and 3q21 chromosome regions including T-cell coreceptor and ligand genes may influence susceptibility to pemphigus foliaceus, Hum. Immunol. 71 (2010) 809–817. https://doi.org/10.1016/j.humimm.2010.04.001.

[20] D.G. Augusto, S.C. Lobo-Alves, M.F. Melo, N.F. Pereira, M.L. Petzl-Erler, Activating KIR and HLA Bw4 ligands are associated to decreased susceptibility to pemphigus foliaceus, an autoimmune blistering skin disease, PLoS One. 7 (2012). https://doi.org/10.1371/journal.pone.0039991.

[21] B.Z. Piovezan, M.L. Petzl-Erler, Both qualitative and quantitative genetic variation of MHC class II molecules may influence susceptibility to autoimmune diseases: the case of endemic pemphigus foliaceus., Hum. Immunol. 74 (2013) 1134–40. https://doi.org/10.1016/j.humimm.2013.06.008.

[22] M.J.F. Brochado, D.F. Nascimento, W. Campos, N.H.S. Deghaide, E.A. Donadi, A.M. Roselino, Differential HLA class I and class II associations in pemphigus foliaceus and pemphigus vulgaris patients from a prevalent Southeastern Brazilian region., J. Autoimmun. 72 (2016) 19–24. https://doi.org/10.1016/j.jaut.2016.04.007.

[23] C.M. Camargo, D.G. Augusto, M.L. Petzl-Erler, Differential gene expression levels might explain association of LAIR2 polymorphisms with pemphigus., Hum. Genet. 135 (2016) 233–44. https://doi.org/10.1007/s00439-015-1626-6.

[24] A. Salviano-Silva, M.L. Petzl-Erler, A.B.W. Boldt, CD59 polymorphisms are associated with gene expression and different sexual susceptibility to pemphigus foliaceus., Autoimmunity. 50 (2017) 377–385. https://doi.org/10.1080/08916934.2017.1329830.

[25] V. Bumiller-Bini, G.A. Cipolla, R.C. de Almeida, M.L. Petzl-Erler, D.G. Augusto, A.B.W. Boldt, Sparking Fire Under the Skin? Answers From the Association of Complement Genes With Pemphigus Foliaceus., Front. Immunol. 9 (2018) 695. https://doi.org/10.3389/fimmu.2018.00695.

[26] T.D.J. Farias, D.G. Augusto, R.C. de Almeida, D. Malheiros, M.L. Petzl-Erler, Screening the full leucocyte receptor complex genomic region revealed associations with pemphigus that might be explained by gene regulation., Immunology. 156 (2019) 86–93. https://doi.org/10.1111/imm.13003.

[27] S.C. Lobo-Alves, D.G. Augusto, W.C.S. Magalhães, E. Tarazona-Santos, M.F. Lima-Costa, M.L. Barreto, B.L. Horta, R.C. de Almeida, M.L. Petzl-Erler, Long noncoding RNA polymorphisms influence susceptibility to endemic pemphigus foliaceus, Br. J. Dermatol. (2019). https://doi.org/10.1111/bjd.17640.

[28] M.E. Dinger, P.P. Amara, T.R. Mercer, K.C. Pang, S.J. Bruce, B.B. Gardiner, M.E. Askarian-Amiri, K. Ru, G. Soldà, C. Simons, S.M. Sunkin, M.L. Crowe, S.M. Grimmond, A.C. Perkins, J.S. Mattick, Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation, Genome Res. 18 (2008) 1433–1445. https://doi.org/10.1101/gr.078378.108.

[29] T.R. Mercer, M.E. Dinger, S.M. Sunkin, M.F. Mehler, J.S. Mattick, Specific expression of long noncoding RNAs in the mouse brain, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 716–721. https://doi.org/10.1073/pnas.0706729105.

[30] A. Salviano-Silva, S.C. Lobo-Alves, R.C. de Almeida, D. Malheiros, M.L. Petzl-Erler, Besides Pathology: Long Non-Coding RNA in Cell and Tissue Homeostasis., Non-Coding RNA. 4 (2018). https://doi.org/10.3390/ncrna4010003.

[31] G.A. Cipolla, J.C. de Oliveira, A. Salviano-Silva, S.C. Lobo-Alves, D.S. Lemos, L.C. Oliveira, T.S. Jucoski, C. Mathias, G.A. Pedroso, E.P. Zambalde, D.F. Gradia, Long non-coding RNAs in multifactorial diseases: Another layer of complexity, Non-Coding RNA. 4 (2018). https://doi.org/10.3390/ncrna4020013.

[32] J.C. de Oliveira, L.C. Oliveira, C. Mathias, G.A. Pedroso, D.S. Lemos, A. Salviano-Silva, T.S. Jucoski, S.C. Lobo-Alves, E.P. Zambalde, G.A. Cipolla, D.F. Gradia, Long non-coding RNAs in cancer: Another layer of complexity, J. Gene Med. 21 (2019). https://doi.org/10.1002/jgm.3065.

[33] B. Hrdlickova, V. Kumar, K. Kanduri, D. V Zhernakova, S. Tripathi, J. Karjalainen, R.J. Lund, Y. Li, U. Ullah, R. Modderman, W. Abdulahad, H. Lähdesmäki, L. Franke, R. Lahesmaa, C. Wijmenga, S. Withoff, Expression profiles of long non-coding RNAs located in autoimmune disease-associated regions reveal immune cell-type specificity., Genome Med. 6 (2014) 88. https://doi.org/10.1186/s13073-014-0088-0.

[34] G.A. Cipolla, J.K. Park, L.A. de Oliveira, S.C. Lobo-Alves, R.C. de Almeida, T.D.J. Farias, D. de S. Lemos, D. Malheiros, R.M. Lavker, M.L. Petzl-Erler, A 3'UTR polymorphism marks differential KLRG1 mRNA levels through disruption of a miR-584-5p binding site and associates with pemphigus foliaceus susceptibility., Biochim. Biophys. Acta. 1859 (2016) 1306–13. https://doi.org/10.1016/j.bbagrm.2016.07.006.

[35] D.W. Russell, J. Sambrook, Molecular cloning: A laboratory manual, Chapter 5, 2001.

[36] E. Schmidt, C. Dähnrich, A. Rosemann, C. Probst, L. Komorowski, S. Saschenbrecker, W. Schlumberger, W. Stöcker, T. Hashimoto, E.-B. Bröcker, A. Recke, C. Rose, D. Zillikens, Novel ELISA systems for antibodies to desmoglein 1 and 3: correlation of disease activity with serum autoantibody levels in individual pemphigus patients., Exp. Dermatol. 19 (2010) 458–63. https://doi.org/10.1111/j.1600-0625.2010.01069.x.

[37] E. Schmidt, M. Goebeler, M. Hertl, M. Sárdy, C. Sitaru, R. Eming, S.C. Hofmann, N. Hunzelmann, J.S. Kern, H. Kramer, H.-D. Orzechowski, C. Pfeiffer, V. Schuster, B. Sporbeck, M. Sticherling, M. Worm, D. Zillikens, A. Nast, S2k guideline for the diagnosis of pemphigus vulgaris/foliaceus and bullous pemphigoid., J. Dtsch. Dermatol. Ges. 13 (2015) 713–27. https://doi.org/10.1111/ddg.12612.

[38] P.J. Volders, J. Anckaert, K. Verheggen, J. Nuytens, L. Martens, P. Mestdagh, J. Vandesompele, Lncipedia 5: Towards a reference set of human long non-coding rnas, Nucleic Acids Res. 47 (2019) D135–D139. https://doi.org/10.1093/nar/gky1031.

[39] C.C. Chang, C.C. Chow, L.C.A.M. Tellier, S. Vattikuti, S.M. Purcell, J.J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets, Gigascience. 4 (2015). https://doi.org/10.1186/s13742-015-0047-8.

[40] M.J. Machiela, S.J. Chanock, LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants., Bioinformatics. 31 (2015) 3555–7. https://doi.org/10.1093/bioinformatics/btv402.

[41] W. McLaren, L. Gil, S.E. Hunt, H.S. Riat, G.R.S. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The Ensembl Variant Effect Predictor, Genome Biol. 17 (2016). https://doi.org/10.1186/s13059-016-0974-4.

[42] F.-O. Desmet, D. Hamroun, M. Lalande, G. Collod-Béroud, M. Claustres, C. Béroud, Human Splicing Finder: an online bioinformatics tool to predict splicing signals., Nucleic Acids Res. 37 (2009) e67. https://doi.org/10.1093/nar/gkp215.

[43] S.M. Wu, H. Liu, P.J. Huang, I.Y.F. Chang, C.C. Lee, C.Y. Yang, W.S. Tsai, B.C.M. Tan, circlncRNAnet: An integrated web-based resource for mapping functional networks of long or circular forms of noncoding RNAs, Gigascience. 7 (2018) 1–10. https://doi.org/10.1093/gigascience/gix118.

[44] The Cancer Genome Atlas Program - National Cancer Institute, (n.d.). https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga (accessed September 17, 2019).

[45] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method, Methods. 25 (2001) 402–408. https://doi.org/10.1006/meth.2001.1262.

[46] Z. Liu, E. Dang, B. Li, H. Qiao, L. Jin, J. Zhang, G. Wang, Dysfunction of CD19+CD24hiCD27+ B regulatory cells in patients with bullous pemphigoid, Sci. Rep. 8 (2018). https://doi.org/10.1038/s41598-018-19226-z.

[47] A. Frezzolini, G. Cianchini, M. Ruffelli, S. Cadoni, P. Puddu, O. De Pità, Interleukin-16 expression and release in bullous pemphigoid, Clin. Exp. Immunol. 137 (2004) 595–600. https://doi.org/10.1111/j.1365-2249.2004.02570.x.

[48] D. Purzycka-Bohdan, A. Szczerkowska-Dobosz, M. Zablotna, J. Wierzbicka, A. Piotrowska, M.A. Zmijewski, B. Nedoszytko, R. Nowicki, Assessment of interleukin 16 serum levels and skin expression in psoriasis patients in correlation with clinical severity of the disease, PLoS One. 11 (2016). https://doi.org/10.1371/journal.pone.0165577.

[49] Y.-C. Tsai, T.-F. Tsai, Anti-interleukin and interleukin therapies for psoriasis: current evidence and clinical usefulness., Ther. Adv. Musculoskelet. Dis. 9 (2017) 277–294. https://doi.org/10.1177/1759720X17735756.

[50] Z. Abu-Zahab, N.M. Gad, S. Sabry, S.A. Nassib, CD81 and CD19 as Marker of Activity in Systemic Lupus Erythematosus Disease., Egypt. J. Immunol. 24 (2017) 141–149. http://www.ncbi.nlm.nih.gov/pubmed/29528587 (accessed September 17, 2019).

[51] L.R. Lard, B.O. Roep, C.A. Verburgh, A.H. Zwinderman, T.W.J. Huizinga, Elevated IL-16 levels in patients with systemic lupus erythematosus are associated with disease severity but not with genetic susceptibility to lupus., Lupus. 11 (2002) 181–5. https://doi.org/10.1191/0961203302lu176sr.

[52] H. Xue, L. Gao, Y. Wu, W. Fang, L. Wang, C. Li, Y. Li, W. Liang, L. Zhang, The IL-16 gene polymorphisms and the risk of the systemic lupus erythematosus., Clin. Chim. Acta. 403 (2009) 223–5. https://doi.org/10.1016/j.cca.2009.03.016.

[53] U. Blank, I.K. Madera-Salcedo, L. Danelli, J. Claver, N. Tiwari, E. Sánchez-Miranda, G. Vázquez-Victorio, K.A. Ramírez-Valadez, M. Macias-Silva, C. González-Espinosa, Vesicular trafficking and signaling for cytokine and chemokine secretion in mast cells., Front. Immunol. 5 (2014) 453. https://doi.org/10.3389/fimmu.2014.00453.

[54] J. Bentham, D.L. Morris, D.S.C. Graham, C.L. Pinder, P. Tombleson, T.W. Behrens, J. Martín, B.P. Fairfax, J.C. Knight, L. Chen, J. Replogle, A.-C. Syvänen, L. Rönnblom, R.R. Graham, J.E. Wither, J.D. Rioux, M.E. Alarcón-Riquelme, T.J. Vyse, Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus., Nat. Genet. 47 (2015) 1457–1464. https://doi.org/10.1038/ng.3434.

[55] G.R. Juszczak, A.M. Stankiewicz, Glucocorticoids, genes and brain function., Prog. Neuropsychopharmacol. Biol. Psychiatry. 82 (2018) 136–168. https://doi.org/10.1016/j.pnpbp.2017.11.020.

[56] C.D. Wells, J.P. Fawcett, A. Traweger, Y. Yamanaka, M. Goudreault, K. Elder, S. Kulkarni, G. Gish, C. Virag, C. Lim, K. Colwill, A. Starostine, P. Metalnikov, T. Pawson, A Rich1/Amot Complex Regulates the Cdc42 GTPase and Apical-Polarity Proteins in Epithelial Cells, Cell. 125 (2006) 535–548. https://doi.org/10.1016/j.cell.2006.02.045.

## *Figure Captions*

**Graphical Abstract.** Genetic associations of immune-related lncRNAs in pemphigus foliaceus.

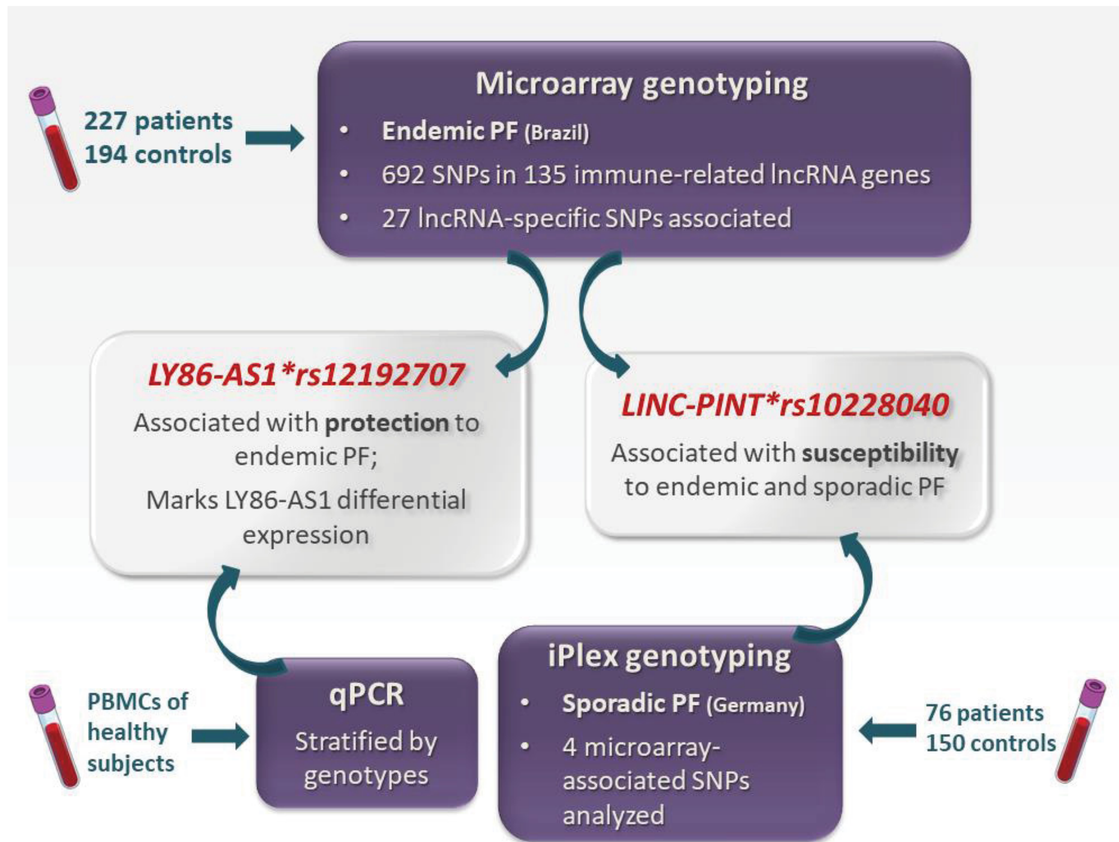
The *LY86-AS1* SNP *rs12192707*A* is associated with protection against endemic pemphigus foliaceus and with lower levels of LY86-AS1 lncRNA. On the other hand, the *LINC-PINT* SNP *rs10228040*A* is associated with increased susceptibility for both endemic and sporadic forms of the disease.

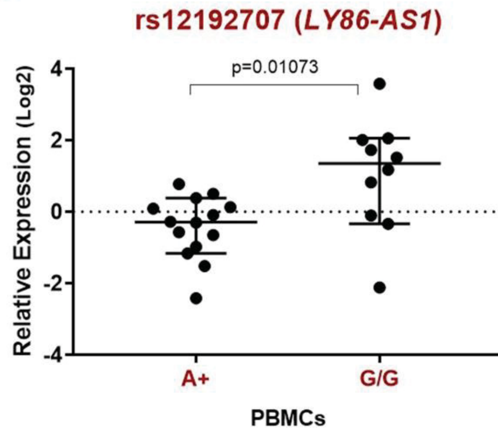
**Figure 1.** Expression RNA levels of *LY86-AS1* and *LINC-PINT* genes stratified by genotypes.

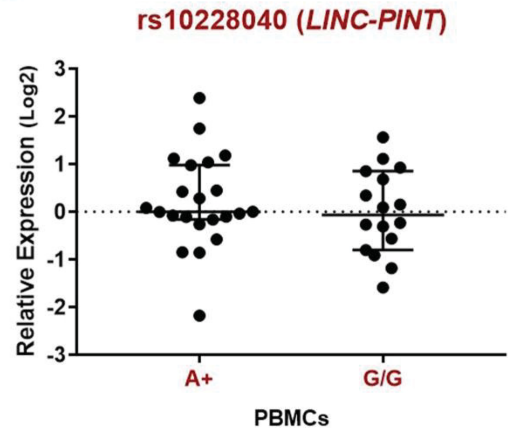
Relative RNA levels of (A) *LY86-AS1* and (B) *LINC-PINT* genes in healthy controls´ PBMC stratified according A>G genotypes. The presence of rs12192707**A* allele (*A/A* and *A/G* genotypes) is associated with lower expression of *LY86-AS1*. No difference in the expression level was observed for the *LINC-PINT* genotypes. Fold-change values were calculated through the $2^{-\Delta\Delta Ct}$ method and normalized by Log2. The horizontal bars at the scatter plots indicate the median. *P* values indicate a statistical significance at the 0.05 level and were calculated with Mann–Whitney's test.

**Supp. Figure 1.** Gene ontology and enriched pathways of genes coexpressed with associated lncRNAs.

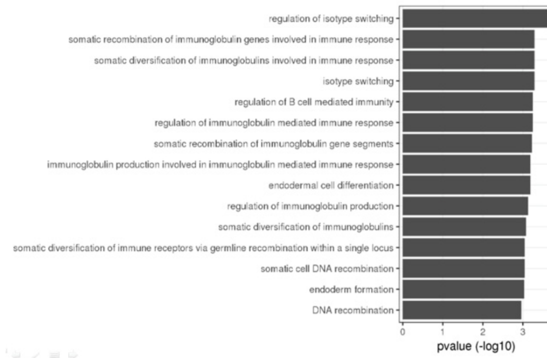
Source: circlncRNAnet database.


227 patients
194 controls
Microarray genotyping
Endemic PF (Brazil)
692 SNPs in 135 immune-related lncRNA genes
27 lncRNA-specific SNPs associated
LY86-AS1*rs12192707
Associated with protection to endemic PF;
Marks LY86-AS1 differential expression
LINC-PINT*rs10228040
Associated with susceptibility to endemic and sporadic PF
PBMCs of healthy subjects
qPCR
Stratified by genotypes
iPlex genotyping
Sporadic PF (Germany)
4 microarray-associated SNPs analyzed
76 patients
150 controls

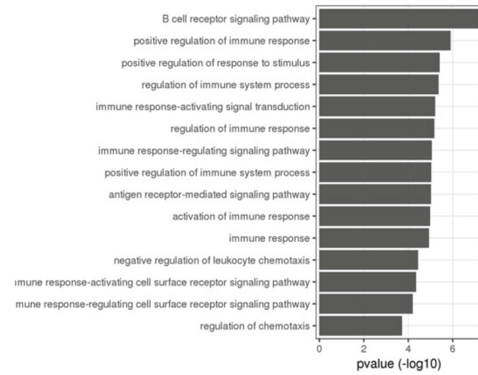

A
rs12192707 (LY86-AS1)
p=0.01073
Relative Expression (Log2)
4
2
0
-2
-4
A+
G/G
PBMCs


B
rs10228040 (LINC-PINT)
Relative Expression (Log2)
3
2
1
0
-1
-2
-3
A+
G/G
PBMCs

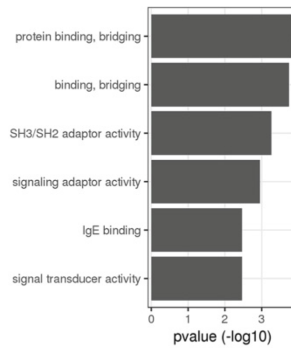
**A** **Enriched biological processes (GO) – CASC15**



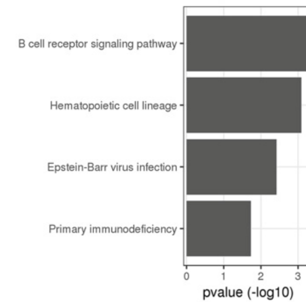

**B** **Enriched biological processes (GO) – LY86-AS1**




**C** **Enriched molecular functions (GO) – LY86-AS1**




**D** **Enriched pathways (KEGG) – LY86-AS1**



## *TABLES*

**Table 1.** Specific primers used for qPCR.

| Gene | Forward Primer 5'-3' | Reverse Primer 5'-3' | Size |
|---|---|---|---|
| *LINC-PINT* | ACCACTGAACAGGAAAAATGAGG | TACCTCATCTGCGAGGAGACA | 131 nt |
| *LY86-AS1* | CGAACAGAGTCAAGTGGAATCAAAG | GCCCAAGAATCAACAGGTAATGTC | 49 nt |
| *GAPDH* | AGGGCTGCTTTTAACTCTGGT | CCCCACTTGATTTTGGAGGGA | 206 nt |

5'-3' sequences of forward and reverse specific primers used for lncRNAs quantification, and the amplicon size. LINC-PINT - lincRNA p53 Induced Transcript; LY86-AS1 – Lymphocyte Antigen 86 Antisense RNA 1; GAPDH - Glyceraldehyde-3-Phosphate Dehydrogenase; nt - nucleotides.

**Table 2.** Endemic pemphigus foliaceus-associated SNPs located in lncRNA genes that do not overlap with protein-coding genes, at the additive model.

| SNP | Genomic position (hg19) | lncRNA genes | Gene class | OR | L95 | U95 | *p*-value |
|---|---|---|---|---|---|---|---|
| rs2473122*A | chr6:22312287 | *CASC15* | Antisense | 1.721 | 1.138 | 2.601 | 0.01 |
| rs10228040*A | chr7:130653616 | *LINC-PINT* | Antisense | 1.465 | 1.089 | 1.971 | 0.0116 |
| rs7812207*A | chr7:130784051 | *LINC-PINT* | Antisense | 0.614 | 0.418 | 0.902 | 0.0128 |
| rs12192707*A | chr6:6479161 | *LY86-AS1* | Antisense | 0.68 | 0.501 | 0.924 | 0.0136 |
| rs4455715*A | chr6:6410663 | *LY86-AS1* | Antisense | 1.425 | 1.066 | 1.906 | 0.0169 |
| rs742286*A | chr6:22391406 | *CASC15* | Antisense | 1.474 | 1.067 | 2.037 | 0.0186 |
| rs10259462*G | chr7:130706995 | *LINC-PINT* | Antisense | 1.518 | 1.071 | 2.154 | 0.0192 |
| rs1418358*A | chr9:29132164 | *lnc-LINGO2-5* | Intergenic | 1.418 | 1.056 | 1.905 | 0.0203 |
| rs7847094* | chr9:29024894 | *lnc-LINGO2-5* | Intergenic | 0.713 | 0.5315 | 0.956 | 0.0236 |
| rs674485*G | chr11:65197393 | *NEAT1* | Intergenic | 1.381 | 1.044 | 1.827 | 0.0238 |
| rs7468614*A | chr9:29054853 | *lnc-LINGO2-5* | Intergenic | 1.428 | 1.048 | 1.945 | 0.024 |
| rs1294468*G | chr6:6544702 | *LY86-AS1* | Antisense | 1.375 | 1.042 | 1.816 | 0.0245 |
| rs4712624*G | chr6:21693152 | *CASC15* | Antisense | 1.442 | 1.038 | 2.003 | 0.0289 |
| rs1884537*A | chr14:101251989 | *MEG3* | Antisense | 0.736 | 0.558 | 0.969 | 0.029 |
| rs7767991*A | chr6:22380084 | *CASC15* | Antisense | 0.715 | 0.527 | 0.968 | 0.0303 |
| rs13242887*G | chr7:130751832 | *LINC-PINT* | Antisense | 0.672 | 0.471 | 0.96 | 0.0328 |
| rs498688*G | chr3:187679835 | *LINC01991* | Intergenic | 1.362 | 1.031 | 1.8 | 0.0357 |
| rs12376696*A | chr9:29024733 | *lnc-LINGO2-5* | Intergenic | 1.425 | 1.018 | 1.994 | 0.0388 |
| rs7768473*A | chr6:22431621 | *CASC15* | Antisense | 0.733 | 0.545 | 0.986 | 0.0399 |
| rs7870900*G | chr9:29153469 | *lnc-LINGO2-5* | Intergenic | 1.359 | 1.014 | 1.822 | 0.0404 |
| rs2870957*C | chr12:68486435 | *IFNG-AS1* | Antisense | 0.713 | 0.516 | 0.986 | 0.0408 |
| rs6974804*G | chr7:130783121 | *LINC-PINT* | Antisense | 1.367 | 1.011 | 1.848 | 0.0423 |
| rs10853784*A | chr19:47793185 | *lnc-C5AR1* | Overlapping | 0.7494 | 0.5661 | 0.992 | 0.0438 |
| rs7742855*A | chr6:22044638 | *CASC15* | Antisense | 1.354 | 1.007 | 1.82 | 0.0448 |
| rs217727*A | chr11:2016908 | *H19* | Intergenic | 0.6995 | 0.493 | 0.9924 | 0.0452 |
| rs12659504*G | chr5:148805005 | *lnc-IL17B-5* | Intergenic | 1.533 | 1.008 | 2.333 | 0.046 |
| rs1418344*G | chr9:29084708 | *lnc-LINGO2-5* | Intergenic | 1.342 | 1.002 | 1.797 | 0.0485 |

Chr – chromosome; OR – Odds Ratio; L95 – 95% confidence interval lower limit; U95 – 95% confidence interval upper limit; CASC15 - Cancer Susceptibility 15; LINC-PINT - lincRNA p53 Induced Transcript; LY86-AS1 - Lymphocyte Antigen 86 Antisense RNA 1; lnc-LINGO2-5 - lncRNA Leucine rich repeat and Ig domain containing 2 transcript 5; NEAT1 - Nuclear Paraspeckle Assembly Transcript 1; MEG3 - Maternally Expressed 3; IFNG-AS1 – Interferon-Gamma Antisense RNA1; lnc-C5AR1 - lncRNA Complement component 5a receptor 1; H19 - Imprinted maternally expressed transcript; lnc-IL17B-5 - lncRNA Interleukin 17B transcript 5. Genomic locations according to LNCipedia database.

**Table 3.** Association analysis for *rs10228040*A* (*LINC-PINT*) evaluation in sporadic pemphigus foliaceus.

| | **rs10228040*A (LINC-PINT)** | | |
|---|---|---|---|
| Models | OR | *p* | CI 95% |
| Additive | 2.284 | **0.0021** | 1.35-3.86 |
| Dominant | 2.499 | **0.0055** | 1.31-4.78 |
| Recessive | 3.319 | **0.0484** | 1.0-10.92 |

In bold: significant *p* values. OR – Odds Ratio; CI 95% - 95% confidence interval.

## *Supplementary tables*

**Supp. Table 1. Immune-related lncRNA genes selected for genotyping data extraction.**

| lncRNA gene | Class | Chr | Initial position | Final position | Related to genes of |
|---|---|---|---|---|---|
| GAS5 | Antisense | 1 | 173820423 | 173838144 | Immune-related |
| lnc-PROX1-41 | Intergenic | 1 | 214098092 | 214099997 | Immune-related |
| PACER | Intergenic | 1 | 186649754 | 186654045 | Immune-related |
| lnc-IL6R-1 | Antisense | 1 | 154452488 | 154453977 | PF associations |
| IL6R-AS1 | Antisense | 1 | 154374804 | 154379040 | PF associations |
| lnc-C8A-1 | Antisense | 1 | 57429559 | 57480373 | PF associations |
| lnc-CFH-1 | Intergenic | 1 | 195450556 | 195469368 | PF associations |
| lnc-CFH-4 | Antisense | 1 | 196150517 | 196249218 | PF associations |
| lnc-CFHR5-5 | Sense-overlapping | 1 | 197251885 | 197312487 | PF associations |
| lnc-CFHR5-6 | Sense-intronic | 1 | 197332947 | 197333208 | PF associations |
| lnc-CR1 | Sense-intronic | 1 | 207831799 | 207832487 | PF associations |
| lnc-CR2 -3 | Intergenic | 1 | 207607630 | 207619658 | PF associations |
| lnc-IL23R | Intergenic | 1 | 67743735 | 67744462 | PV associations |
| lnc-IL10-1 | Antisense | 1 | 206869182 | 206869614 | PV associations |
| lnc-IL10-2 | Sense-intronic | 1 | 206776437 | 206781648 | PV associations |
| lnc-IL10-5 | Sense-intronic | 1 | 206940947 | 206945780 | PV associations |
| MIR4435-2HG | Sense-intronic | 2 | 111794353 | 112456610 | Immune-related |
| LINC00487 | Intergenic | 2 | 6868309 | 6910442 | Immune-related |
| CNNM3-DT | Antisense | 2 | 97477562 | 97482303 | Immune-related |
| lnc-IL1B-1 | Intergenic | 2 | 113576774 | 113581878 | Immune-related |
| lnc-IL1B-2 | Intergenic | 2 | 113636769 | 113637743 | Immune-related |
| lnc-PDCD1-1 | Sense-intronic | 2 | 242835752 | 242844846 | PF associations |
| FGD5-AS1 | Antisense | 3 | 14919069 | 14991070 | Immune-related |
| Lnc-CCRL2 | Sense-intronic | 3 | 46451005 | 46454488 | Immune-related |
| LINC00877 | Intergenic | 3 | 72084451 | 72328654 | Immune-related |
| LINC01991 | Intergenic | 3 | 187676548 | 187694195 | Immune-related |
| lnc-MASP1-1 | Antisense | 3 | 186914878 | 186925417 | PF associations |
| lnc-MASP1-2 e 3 | Intergenic | 3 | 187131305 | 187133958 | PF associations |
| lnc-ARPP21-1 | Intergenic | 3 | 35913358 | 35913690 | PV associations |
| lnc-ARPP22-2 | Intergenic | 3 | 34914133 | 34915346 | PV associations |
| ARPP21-AS1 | Antisense | 3 | 35691689 | 35693453 | PV associations |
| lnc-PTPRG-1 | Intergenic | 3 | 61237271 | 61302257 | PV associations |
| lnc-PTPRG-4 | Intergenic | 3 | 61395234 | 61398853 | PV associations |
| lnc-PTPRG-5 | Antisense | 3 | 60602555 | 60603812 | PV associations |
| PTPRG-AS1 | Antisense | 3 | 62242333 | 62355017 | PV associations |
| GPD1L-1 | Intergenic | 3 | 32232239 | 32233119 | PV associations |
| GPD1L-2 | Sense-intronic | 3 | 32303234 | 32305396 | PV associations |
| LINC00989 | Intergenic | 4 | 80413570 | 80497614 | Immune-related |
| SMAD1-AS1 | Antisense | 4 | 146435730 | 146438346 | Immune-related |
| LEF1-AS1 | Intergenic/antisense | 4 | 109092927 | 109226083 | Immune-related |
| lnc-ZNF827-2 | Intergenic | 4 | 146973756 | 146977914 | Immune-related |
| lnc-CLINT1-2 | Intergenic | 5 | 157652360 | 157661384 | Immune-related |

| | | | | | |
|---|---|---|---|---|---|
| TH2LCRR | Antisense | 5 | 131966281 | 131999964 | Immune-related |
| lnc-IL7R-1 | Intergenic | 5 | 35938903 | 35940095 | Immune-related |
| lnc-IL7R-2 | Antisense | 5 | 35974665 | 35975873 | Immune-related |
| lnc-IL4 | Antisense | 5 | 132024373 | 132059417 | PF associations |
| lnc-C9-1 | Sense-intronic | 5 | 39331875 | 39462400 | PF associations |
| lnc-NR3C1-1 | Intergenic | 5 | 142869420 | 142910915 | PV associations |
| lnc-NR3C1-2 | Sense-intronic | 5 | 143543427 | 143550204 | PV associations |
| lnc-NR3C1-3 | Intergenic | 5 | 142621254 | 142624088 | PV associations |
| lnc-IL17B-2 | Sense-intronic | 5 | 148872949 | 148884233 | PV associations |
| lnc-IL17B-3 | Antisense | 5 | 148543518 | 148656368 | PV associations |
| lnc-IL17B-4 | Sense-intronic | 5 | 148750887 | 148753905 | PV associations |
| lnc-IL17B-5 | Intergenic | 5 | 148804756 | 148812399 | PV associations |
| lnc-IL17B-6 | Sense-intronic | 5 | 148924946 | 148925373 | PV associations |
| lnc-IL17B-7 | Sense-intronic | 5 | 148422283 | 148442625 | PV associations |
| lnc-IL17B-8 | Sense-intronic | 5 | 148407948 | 148415584 | PV associations |
| lnc-IL17B-9 | Bidirectional promoter | 5 | 148724344 | 148724906 | PV associations |
| MYB-AS1 | Antisense | 6 | 135514749 | 135557349 | Immune-related |
| PSORS1C3 | Sense-intronic | 6 | 31139386 | 31154249 | Immune-related |
| lnc-BACH2 | Sense-intronic | 6 | 91280770 | 91281081 | Immune-related |
| CASC15 | Antisense | 6 | 21664451 | 22517940 | Immune-related |
| LY86-AS1 | Antisense | 6 | 6346337 | 6623059 | Immune-related |
| lnc-MICB-4 | Intergenic | 6 | 31483756 | 31483988 | PV associations |
| lnc-NOTCH4 | Sense-intronic | 6 | 32162620 | 32164754 | PV associations |
| lnc-BTNL2-1 | Sense-intronic | 6 | 32372815 | 32374907 | PV associations |
| lnc-BTNL2-2 | Intergenic | 6 | 32403381 | 32405137 | PV associations |
| lnc-IL17A-1 | Intergenic | 6 | 52066227 | 52068399 | PV associations |
| lnc-IL17A-2 | Antisense | 6 | 51464521 | 51487682 | PV associations |
| lnc-IL17A-3 | Antisense | 6 | 51840458 | 51840755 | PV associations |
| lnc-IL17A-4 | Sense-intronic | 6 | 52257087 | 52258349 | PV associations |
| lnc-IL17A-5 | Sense-intronic | 6 | 52262356 | 52262946 | PV associations |
| LINC-PINT | Antisense | 7 | 130476023 | 130875188 | Immune-related |
| TP53TG1 | Bidirectional promoter | 7 | 86953598 | 86974883 | Immune-related |
| HOTAIRM1 | Antisense | 7 | 27135699 | 27139884 | Immune-related |
| lnc-IL6-3 | Antisense | 7 | 22551434 | 22690102 | PF associations |
| SAS-ZFAT | Antisense | 8 | 135610314 | 135612932 | Immune-related |
| lnc-ST18 | Sense-intronic | 8 | 53085073 | 53373519 | PV associations |
| NRON | Antisense | 9 | 129170053 | 129172783 | Immune-related |
| RMRP | Bidirectional promoter | 9 | 35657748 | 35658015 | Immune-related |
| lnc-C5 | Sense-intronic | 9 | 123686736 | 123765819 | PF associations |
| lnc-LINGO2-1 | Intergenic | 9 | 27937615 | 27944495 | PV associations |
| lnc-LINGO2-2 | Intergenic | 9 | 29185807 | 29214176 | PV associations |
| lnc-LINGO2-3 | Sense-intronic | 9 | 27948076 | 27948746 | PV associations |
| lnc-LINGO2-4 | Intergenic | 9 | 27927621 | 27934495 | PV associations |
| lnc-LINGO2-5 | Intergenic | 9 | 28937769 | 29203552 | PV associations |
| lnc-LINGO2-6 | Intergenic | 9 | 29254073 | 29254424 | PV associations |

| | | | | | |
|---|---|---|---|---|---|
| lnc-LINGO2-7 | Intergenic | 9 | 29824740 | 29826709 | PV associations |
| GATA3-AS1 | Antisense/intergenic | 10 | 8058532 | 8096328 | Immune-related |
| lnc-THNSL1-3 | Sense-intronic | 10 | 24536051 | 24544975 | Immune-related |
| LINC00678 | Intergenic | 11 | 27620891 | 27656267 | Immune-related |
| lnc-ZC3H12C | Antisense | 11 | 109731130 | 110168471 | Immune-related |
| NCAM1-AS1 | Antisense | 11 | 113133455 | 113144798 | Immune-related |
| H19 | Intergenic | 11 | 2016360 | 2022940 | Immune-related |
| MALAT1 | Intergenic | 11 | 65263738 | 65276556 | Immune-related |
| NEAT1 | Intergenic | 11 | 65184053 | 65217564 | Immune-related |
| lnc-SPI1 | Antisense | 11 | 47404699 | 47430741 | Immune-related |
| lnc-CD59-1 | Sense-intronic | 11 | 33720004 | 33721943 | PF associations |
| lnc-CD59-2 | Sense-intronic | 11 | 33767483 | 33770358 | PF associations |
| HOTAIR | Antisense | 12 | 54356092 | 54368740 | Immune-related |
| IFNG-AS1 | Antisense | 12 | 68383162 | 68628466 | Immune-related |
| THRIL | Antisense | 12 | 125509980 | 125513897 | Immune-related |
| NRAV | Antisense | 12 | 120918922 | 120933813 | Immune-related |
| lnc-IL23A-2 | Antisense | 12 | 56701427 | 56703233 | PV associations |
| lnc-IL23A-4 | Antisense | 12 | 56754022 | 56754942 | PV associations |
| lnc-IL23A-5 | Antisense | 12 | 56709889 | 56710468 | PV associations |
| lnc-IL23A-6 | Antisense | 12 | 56694177 | 56708592 | PV associations |
| lnc-IL17D-3 | Sense-intronic | 13 | 21220473 | 21224718 | PV associations |
| lnc-IL17D-4 | Antisense | 13 | 21347417 | 21349303 | PV associations |
| lnc-BRF1-24 | Intergenic | 14 | 106714360 | 106714821 | Immune-related |
| MEG3 | Antisense/Intergenic | 14 | 101245747 | 101327368 | Immune-related |
| MAFTRR | Intergenic | 16 | 79691311 | 79804827 | Immune-related |
| lnc-CIITA-4 | Sense-intronic | 16 | 11150413 | 11150891 | PF associations |
| lnc-CD19-1 | Antisense | 16 | 28939983 | 28940670 | PF associations |
| lnc-CD19-2 | Antisense | 16 | 28936534 | 28938053 | PF associations |
| lnc-ITGAX-5 | Intergenic | 16 | 31398262 | 31399205 | PF associations |
| lnc-ITGAX-4 | Intergenic | 16 | 31361892 | 31363822 | PF associations |
| lnc-DC | Intergenic | 17 | 58160924 | 58169300 | Immune-related |
| TCF4-AS1 | Antisense | 18 | 53119764 | 53150171 | PV associations |
| TCF4-AS2 | Antisense | 18 | 53159406 | 53163623 | PV associations |
| lnc-TCF4-2 | Sense-intronic | 18 | 54266299 | 54268217 | PV associations |
| lnc-TCF4-6 | Sense-intronic | 18 | 52596107 | 52604508 | PV associations |
| lnc-TCF4-7 | Intergenic | 18 | 53746625 | 53746825 | PV associations |
| lnc-TCF4-8 | Intergenic | 18 | 53440548 | 53447389 | PV associations |
| lnc-KIR2DL3 | Sense-overlapping/intergenic | 19 | 55215730 | 55261861 | Immune-related |
| lnc-KIR3DX1-2 | Intergenic | 19 | 55063253 | 55063539 | Immune-related |
| lnc-C3-1 | Sense-intronic | 19 | 6730066 | 6732079 | PF associations |
| lnc-C3-2 | Sense-intronic | 19 | 6735633 | 6737284 | PF associations |
| lnc-C5AR1 | Sense-overlapping | 19 | 47593166 | 47848011 | PF associations |
| CEBPB-AS1 | Antisense | 20 | 48801135 | 48808606 | Immune-related |
| NORAD | Intergenic | 20 | 34633221 | 34638938 | Immune-related |
| NKILA | Antisense | 20 | 56285239 | 56287836 | Immune-related |

| | | | | | |
|---|---|---|---|---|---|
| lnc-CD40-2 | Antisense | 20 | 44828955 | 44835850 | PF associations |
| lnc-CD40-3 | Antisense | 20 | 44993170 | 45019524 | PF associations |
| ITGB2-AS1 | Antisense | 21 | 46340950 | 46349595 | Immune-related |

**Supp. Table 2. Genomic positions of the analyzed SNPs at immune-related lncRNA genes**

| Chr | Position (hg19) | A1 | A2 |
|---|---|---|---|
| 1 | 57441363 | G | A |
| 1 | 67740835 | C | A |
| 1 | 67747415 | C | A |
| 1 | 154369252 | G | A |
| 1 | 154400015 | C | A |
| 1 | 154438084 | G | A |
| 1 | 154457855 | A | C |
| 1 | 186659125 | A | G |
| 1 | 196170439 | G | A |
| 1 | 196234932 | G | A |
| 1 | 196243012 | G | A |
| 1 | 197257090 | G | A |
| 1 | 197266535 | A | G |
| 1 | 197293463 | G | A |
| 1 | 197306099 | G | A |
| 1 | 197329535 | A | C |
| 1 | 197359999 | G | A |
| 1 | 206776460 | A | G |
| 1 | 206857407 | G | A |
| 1 | 206872487 | G | A |
| 1 | 206943968 | A | C |
| 1 | 206944645 | A | G |
| 1 | 207595822 | G | A |
| 1 | 207840776 | A | G |
| 1 | 214099716 | A | G |
| 2 | 6886617 | G | A |
| 2 | 6891670 | G | A |
| 2 | 111795509 | A | G |
| 2 | 111796833 | G | A |
| 2 | 111797458 | G | A |
| 2 | 111797531 | G | A |
| 2 | 111812617 | A | G |
| 2 | 111813085 | C | A |
| 2 | 111836538 | G | A |
| 2 | 111849659 | G | A |
| 2 | 111850515 | A | G |
| 2 | 111864915 | A | G |
| 2 | 111868010 | A | C |
| 2 | 111875388 | A | G |
| 2 | 111908262 | G | A |

| | | | |
|---|---|---|---|
| 2 | 111918472 | A | G |
| 2 | 111948809 | A | C |
| 2 | 111949327 | G | C |
| 2 | 111957894 | A | G |
| 2 | 111966222 | A | G |
| 2 | 111966540 | G | A |
| 2 | 111976515 | A | G |
| 2 | 111988016 | A | G |
| 2 | 111989372 | C | A |
| 2 | 111992975 | A | G |
| 2 | 112004205 | A | G |
| 2 | 112010486 | G | A |
| 2 | 112189642 | A | G |
| 2 | 112253302 | C | A |
| 2 | 112277492 | C | A |
| 2 | 112409842 | G | A |
| 2 | 112411416 | G | A |
| 2 | 112415726 | A | G |
| 2 | 112428409 | G | A |
| 2 | 112444459 | A | C |
| 2 | 112447888 | A | G |
| 2 | 113570581 | C | A |
| 2 | 113590467 | A | G |
| 2 | 113610780 | G | A |
| 2 | 113641631 | A | G |
| 2 | 242824974 | A | G |
| 2 | 242873800 | A | G |
| 3 | 14923230 | A | G |
| 3 | 14923396 | A | G |
| 3 | 14932156 | A | G |
| 3 | 14939088 | A | G |
| 3 | 14939471 | A | G |
| 3 | 14939479 | C | A |
| 3 | 14941152 | A | C |
| 3 | 14942024 | G | A |
| 3 | 14960204 | A | C |
| 3 | 14964296 | A | G |
| 3 | 14965466 | G | A |
| 3 | 14969197 | A | G |
| 3 | 14981262 | C | A |
| 3 | 32225326 | A | G |
| 3 | 32239401 | A | G |
| 3 | 32304103 | A | C |
| 3 | 34908824 | G | A |
| 3 | 35692163 | C | A |

| | | | |
|---|---|---|---|
| 3 | 35902885 | G | A |
| 3 | 35926289 | A | C |
| 3 | 46451680 | A | G |
| 3 | 60597001 | G | A |
| 3 | 61241178 | C | A |
| 3 | 61252469 | A | G |
| 3 | 61253495 | A | G |
| 3 | 61262149 | A | G |
| 3 | 61273542 | A | G |
| 3 | 61289533 | A | C |
| 3 | 61348592 | A | G |
| 3 | 61406345 | G | A |
| 3 | 62244570 | G | A |
| 3 | 62246422 | G | A |
| 3 | 62272627 | G | A |
| 3 | 62315312 | G | A |
| 3 | 62317005 | G | A |
| 3 | 62319048 | G | A |
| 3 | 62341044 | A | G |
| 3 | 62344960 | A | G |
| 3 | 72097527 | G | A |
| 3 | 72118981 | A | G |
| 3 | 72126377 | G | A |
| 3 | 72126831 | A | C |
| 3 | 72141723 | A | G |
| 3 | 72146520 | A | G |
| 3 | 72150727 | A | C |
| 3 | 72151732 | C | A |
| 3 | 72158910 | G | A |
| 3 | 72161259 | G | A |
| 3 | 72163161 | A | G |
| 3 | 72163948 | G | A |
| 3 | 72170887 | G | A |
| 3 | 72181219 | A | G |
| 3 | 72200540 | C | A |
| 3 | 72213025 | A | G |
| 3 | 72216115 | G | A |
| 3 | 72222004 | G | A |
| 3 | 72222082 | G | A |
| 3 | 72227868 | A | G |
| 3 | 72229196 | C | A |
| 3 | 72231202 | G | A |
| 3 | 72236434 | G | A |
| 3 | 72250823 | G | A |
| 3 | 72266311 | A | G |

| | | | |
|---|---|---|---|
| 3 | 72282932 | A | G |
| 3 | 72285612 | G | A |
| 3 | 72288745 | A | G |
| 3 | 72295961 | A | C |
| 3 | 72300561 | A | G |
| 3 | 72310765 | A | G |
| 3 | 72313715 | A | G |
| 3 | 72315956 | A | G |
| 3 | 72316559 | G | A |
| 3 | 72316954 | G | A |
| 3 | 72317470 | A | G |
| 3 | 72326691 | G | A |
| 3 | 186917586 | A | G |
| 3 | 186917751 | C | G |
| 3 | 186923619 | C | A |
| 3 | 186924035 | G | A |
| 3 | 187128567 | A | G |
| 3 | 187160825 | A | C |
| 3 | 187679835 | G | A |
| 3 | 187688058 | G | A |
| 3 | 187689617 | A | G |
| 3 | 187690288 | G | A |
| 4 | 80416491 | G | A |
| 4 | 80425028 | A | G |
| 4 | 80461045 | A | G |
| 4 | 80482400 | G | A |
| 4 | 109111726 | C | A |
| 4 | 109142299 | C | A |
| 4 | 109163035 | G | A |
| 4 | 109176161 | A | G |
| 4 | 109192876 | A | G |
| 4 | 109196994 | A | G |
| 4 | 109202036 | C | A |
| 4 | 109218677 | A | G |
| 4 | 146974543 | G | A |
| 5 | 35933518 | C | A |
| 5 | 35940365 | A | G |
| 5 | 35973875 | A | C |
| 5 | 35981184 | G | A |
| 5 | 39340135 | A | C |
| 5 | 39349825 | C | A |
| 5 | 39354069 | A | G |
| 5 | 39354542 | A | C |
| 5 | 39357867 | A | G |
| 5 | 39364554 | A | G |

| | | | |
|---|---|---|---|
| 5 | 39370907 | A | G |
| 5 | 39376988 | A | G |
| 5 | 39377132 | A | G |
| 5 | 39394989 | C | A |
| 5 | 39395397 | A | G |
| 5 | 39397132 | T | A |
| 5 | 39415860 | G | A |
| 5 | 39417846 | G | A |
| 5 | 39428050 | A | G |
| 5 | 39434349 | A | G |
| 5 | 39439306 | G | A |
| 5 | 39444239 | G | A |
| 5 | 39451901 | C | A |
| 5 | 131973177 | G | A |
| 5 | 131986463 | C | A |
| 5 | 131995843 | A | G |
| 5 | 131995964 | A | G |
| 5 | 132032369 | A | G |
| 5 | 132040069 | A | G |
| 5 | 132042146 | A | G |
| 5 | 132049027 | G | A |
| 5 | 142612488 | A | G |
| 5 | 142657212 | G | A |
| 5 | 142870815 | A | G |
| 5 | 142873977 | C | A |
| 5 | 142881689 | A | G |
| 5 | 142883695 | G | A |
| 5 | 142895811 | G | A |
| 5 | 142900260 | C | A |
| 5 | 142904601 | A | G |
| 5 | 142904949 | C | A |
| 5 | 143545052 | A | G |
| 5 | 148431964 | A | G |
| 5 | 148544809 | A | G |
| 5 | 148544856 | G | A |
| 5 | 148554730 | A | G |
| 5 | 148559909 | G | A |
| 5 | 148576253 | A | G |
| 5 | 148588846 | A | G |
| 5 | 148620535 | A | G |
| 5 | 148631433 | G | A |
| 5 | 148805005 | G | A |
| 5 | 148810746 | G | A |
| 5 | 148917992 | G | A |
| 5 | 148932568 | A | G |

| | | | |
|---|---|---|---|
| 5 | 157652081 | G | A |
| 5 | 157689852 | A | G |
| 6 | 6355950 | G | A |
| 6 | 6356142 | A | C |
| 6 | 6363273 | A | G |
| 6 | 6366427 | A | G |
| 6 | 6381420 | G | A |
| 6 | 6384148 | A | G |
| 6 | 6397716 | C | A |
| 6 | 6409308 | A | G |
| 6 | 6410663 | A | G |
| 6 | 6412535 | A | G |
| 6 | 6420874 | A | G |
| 6 | 6425334 | A | G |
| 6 | 6432964 | G | A |
| 6 | 6451562 | A | G |
| 6 | 6455153 | C | A |
| 6 | 6461999 | A | G |
| 6 | 6467872 | A | G |
| 6 | 6474600 | A | G |
| 6 | 6476278 | A | G |
| 6 | 6479161 | A | G |
| 6 | 6486785 | G | A |
| 6 | 6491347 | A | G |
| 6 | 6495522 | A | G |
| 6 | 6523310 | A | G |
| 6 | 6531684 | A | G |
| 6 | 6535234 | A | G |
| 6 | 6535323 | A | G |
| 6 | 6541883 | A | G |
| 6 | 6544702 | G | A |
| 6 | 6550156 | G | A |
| 6 | 6555790 | A | G |
| 6 | 6558761 | G | A |
| 6 | 6559568 | G | A |
| 6 | 6574384 | A | C |
| 6 | 6582123 | G | A |
| 6 | 6588881 | A | C |
| 6 | 6591304 | C | A |
| 6 | 6593049 | A | G |
| 6 | 6594321 | G | A |
| 6 | 6601050 | A | G |
| 6 | 6612467 | G | C |
| 6 | 6614433 | A | G |
| 6 | 6614501 | C | A |

| | | | |
|---|---|---|---|
| 6 | 6618373 | G | A |
| 6 | 6618549 | C | A |
| 6 | 6619698 | A | C |
| 6 | 21670816 | A | G |
| 6 | 21691009 | A | G |
| 6 | 21693152 | G | A |
| 6 | 21697176 | G | A |
| 6 | 21701841 | A | G |
| 6 | 21727148 | A | G |
| 6 | 21728317 | A | G |
| 6 | 21729682 | A | G |
| 6 | 21734750 | A | G |
| 6 | 21740321 | G | A |
| 6 | 21753645 | A | C |
| 6 | 21766998 | A | C |
| 6 | 21780789 | G | A |
| 6 | 21781952 | G | A |
| 6 | 21786171 | A | G |
| 6 | 21789088 | A | G |
| 6 | 21796241 | A | G |
| 6 | 21805714 | G | A |
| 6 | 21813153 | G | A |
| 6 | 21815916 | G | A |
| 6 | 21824770 | G | A |
| 6 | 21832546 | G | A |
| 6 | 21834059 | C | A |
| 6 | 21836272 | G | A |
| 6 | 21844709 | A | C |
| 6 | 21847532 | G | A |
| 6 | 21856821 | C | A |
| 6 | 21863899 | A | G |
| 6 | 21874359 | A | C |
| 6 | 21891605 | G | A |
| 6 | 21894448 | A | G |
| 6 | 21903533 | G | A |
| 6 | 21911129 | G | A |
| 6 | 21912632 | G | A |
| 6 | 21921026 | G | A |
| 6 | 21922708 | A | C |
| 6 | 21925491 | A | G |
| 6 | 21929886 | G | A |
| 6 | 21934671 | A | G |
| 6 | 21936868 | G | A |
| 6 | 21943942 | A | G |
| 6 | 21956404 | G | A |

| | | | |
|---|---|---|---|
| 6 | 21969495 | A | G |
| 6 | 21972800 | C | A |
| 6 | 21982253 | A | G |
| 6 | 21985247 | A | G |
| 6 | 21986847 | A | C |
| 6 | 21996860 | A | G |
| 6 | 22002897 | G | A |
| 6 | 22003895 | C | A |
| 6 | 22012566 | C | A |
| 6 | 22017738 | G | A |
| 6 | 22018581 | A | G |
| 6 | 22018994 | G | A |
| 6 | 22044638 | A | G |
| 6 | 22064639 | G | A |
| 6 | 22073654 | A | G |
| 6 | 22075262 | A | G |
| 6 | 22100367 | A | G |
| 6 | 22112043 | G | A |
| 6 | 22122568 | G | A |
| 6 | 22131929 | C | A |
| 6 | 22136409 | G | A |
| 6 | 22140004 | A | G |
| 6 | 22141596 | A | G |
| 6 | 22162129 | G | A |
| 6 | 22166552 | G | A |
| 6 | 22187889 | A | C |
| 6 | 22190338 | G | A |
| 6 | 22194163 | C | A |
| 6 | 22194664 | C | A |
| 6 | 22207220 | G | A |
| 6 | 22207494 | A | G |
| 6 | 22220860 | A | G |
| 6 | 22236978 | C | A |
| 6 | 22238088 | A | G |
| 6 | 22246604 | G | A |
| 6 | 22255514 | G | A |
| 6 | 22272753 | A | G |
| 6 | 22274766 | A | G |
| 6 | 22280214 | C | A |
| 6 | 22281258 | A | G |
| 6 | 22291367 | A | G |
| 6 | 22304204 | A | C |
| 6 | 22307725 | A | C |
| 6 | 22322004 | A | G |
| 6 | 22323312 | G | A |

| | | | |
|---|---|---|---|
| 6 | 22341469 | G | A |
| 6 | 22343301 | G | A |
| 6 | 22359969 | A | G |
| 6 | 22380084 | A | G |
| 6 | 22382907 | G | A |
| 6 | 22390618 | A | G |
| 6 | 22391406 | A | G |
| 6 | 22403373 | A | G |
| 6 | 22404476 | A | G |
| 6 | 22405396 | A | G |
| 6 | 22417829 | A | G |
| 6 | 22423766 | C | A |
| 6 | 22431621 | A | G |
| 6 | 22433910 | A | G |
| 6 | 22446285 | A | C |
| 6 | 22457726 | C | A |
| 6 | 22480823 | A | G |
| 6 | 22484377 | G | A |
| 6 | 22491695 | G | A |
| 6 | 22501947 | G | A |
| 6 | 22504878 | G | A |
| 6 | 22507985 | G | A |
| 6 | 22508420 | G | A |
| 6 | 22515654 | G | A |
| 6 | 22516569 | G | A |
| 6 | 31139452 | A | G |
| 6 | 31140047 | A | G |
| 6 | 31140152 | C | A |
| 6 | 31141127 | G | A |
| 6 | 31141523 | A | G |
| 6 | 31142245 | C | A |
| 6 | 31143582 | A | C |
| 6 | 31143835 | G | C |
| 6 | 31145271 | G | A |
| 6 | 31145920 | A | G |
| 6 | 31147194 | G | A |
| 6 | 31483415 | C | A |
| 6 | 32373312 | A | G |
| 6 | 32373801 | C | A |
| 6 | 32374640 | A | G |
| 6 | 32405026 | A | G |
| 6 | 51465832 | A | G |
| 6 | 51480486 | C | A |
| 6 | 51483961 | A | G |
| 6 | 51484226 | A | G |

| | | | |
|---|---|---|---|
| 6 | 51838263 | G | A |
| 6 | 51873092 | G | A |
| 6 | 52056386 | A | G |
| 6 | 52075816 | A | G |
| 6 | 52263744 | G | A |
| 6 | 91284990 | C | A |
| 6 | 135525396 | A | C |
| 6 | 135535240 | A | G |
| 6 | 135536496 | A | G |
| 6 | 135540380 | G | A |
| 6 | 135541194 | A | G |
| 6 | 135541395 | A | G |
| 6 | 135550343 | A | G |
| 7 | 22556193 | A | G |
| 7 | 22580985 | G | A |
| 7 | 22598742 | G | A |
| 7 | 22601127 | A | C |
| 7 | 22604319 | A | G |
| 7 | 22654968 | A | G |
| 7 | 22663453 | A | C |
| 7 | 27135314 | G | A |
| 7 | 86969559 | G | A |
| 7 | 130480244 | A | G |
| 7 | 130492122 | C | A |
| 7 | 130513816 | G | A |
| 7 | 130550812 | G | A |
| 7 | 130553262 | G | A |
| 7 | 130562824 | G | A |
| 7 | 130566594 | A | G |
| 7 | 130585553 | C | A |
| 7 | 130589791 | A | G |
| 7 | 130616236 | C | A |
| 7 | 130627014 | A | C |
| 7 | 130629901 | A | G |
| 7 | 130641577 | G | A |
| 7 | 130653616 | A | G |
| 7 | 130664881 | A | G |
| 7 | 130667258 | A | G |
| 7 | 130668618 | A | G |
| 7 | 130673153 | G | A |
| 7 | 130678154 | G | A |
| 7 | 130688577 | G | A |
| 7 | 130706995 | G | A |
| 7 | 130717723 | A | G |
| 7 | 130743218 | G | A |

| | | | |
|---|---|---|---|
| 7 | 130751832 | G | A |
| 7 | 130783121 | G | A |
| 7 | 130784051 | A | C |
| 7 | 130813097 | A | C |
| 7 | 130824877 | A | G |
| 7 | 130827699 | A | G |
| 7 | 130839735 | G | A |
| 7 | 130848109 | A | G |
| 7 | 130850903 | G | A |
| 7 | 130854175 | A | G |
| 7 | 130863643 | A | G |
| 8 | 53112386 | G | A |
| 8 | 53125900 | A | G |
| 8 | 53128122 | A | G |
| 8 | 53133166 | G | A |
| 8 | 53141715 | A | G |
| 8 | 53145161 | A | G |
| 8 | 53172806 | G | A |
| 8 | 53175514 | A | C |
| 8 | 53188287 | G | A |
| 8 | 53202655 | A | G |
| 8 | 53202675 | A | G |
| 8 | 53227219 | G | A |
| 8 | 53242775 | A | G |
| 8 | 53259979 | A | C |
| 8 | 53283773 | G | A |
| 8 | 53291161 | G | A |
| 8 | 53295154 | C | A |
| 8 | 53313496 | A | G |
| 8 | 53318644 | A | C |
| 8 | 53320710 | A | G |
| 8 | 53328331 | G | A |
| 8 | 53335677 | G | A |
| 8 | 53341622 | A | G |
| 8 | 135611945 | C | A |
| 8 | 135612595 | A | G |
| 9 | 27931179 | G | A |
| 9 | 27945758 | A | G |
| 9 | 28942538 | A | G |
| 9 | 28959079 | G | A |
| 9 | 28970237 | A | G |
| 9 | 28979255 | G | A |
| 9 | 29003807 | A | G |
| 9 | 29015933 | G | A |
| 9 | 29017258 | A | G |

| | | | |
|---|---|---|---|
| 9 | 29023122 | A | G |
| 9 | 29024733 | A | G |
| 9 | 29024894 | G | A |
| 9 | 29042229 | A | G |
| 9 | 29054853 | A | T |
| 9 | 29084708 | G | A |
| 9 | 29088547 | A | G |
| 9 | 29112301 | A | G |
| 9 | 29119535 | G | A |
| 9 | 29126329 | C | A |
| 9 | 29132164 | A | G |
| 9 | 29153469 | G | A |
| 9 | 29166390 | A | C |
| 9 | 29174711 | G | A |
| 9 | 29192125 | G | A |
| 9 | 29198980 | A | G |
| 9 | 29249698 | G | A |
| 9 | 29261935 | G | A |
| 9 | 29823067 | A | G |
| 9 | 29831400 | A | G |
| 9 | 123690239 | G | A |
| 9 | 123699083 | G | A |
| 9 | 123703894 | A | G |
| 9 | 123705945 | A | T |
| 9 | 123725926 | C | A |
| 9 | 123731408 | A | G |
| 9 | 123757412 | G | A |
| 9 | 129166088 | A | G |
| 9 | 129184309 | A | G |
| 10 | 8074437 | A | C |
| 10 | 8092186 | A | G |
| 10 | 8093034 | C | A |
| 10 | 24537108 | G | A |
| 10 | 24542427 | C | A |
| 10 | 24542941 | A | G |
| 11 | 2016908 | A | G |
| 11 | 2021075 | A | G |
| 11 | 27628412 | G | A |
| 11 | 27635319 | C | A |
| 11 | 27646745 | G | A |
| 11 | 27647285 | A | G |
| 11 | 33709273 | G | A |
| 11 | 47385923 | G | A |
| 11 | 47431529 | A | G |
| 11 | 65190379 | C | A |

| | | | |
|---|---|---|---|
| 11 | 65197393 | G | A |
| 11 | 65203561 | A | G |
| 11 | 65206017 | A | G |
| 11 | 65260646 | G | A |
| 11 | 109747761 | A | G |
| 11 | 109754949 | A | G |
| 11 | 109801094 | G | A |
| 11 | 109805805 | A | G |
| 11 | 109810415 | G | A |
| 11 | 109812865 | A | C |
| 11 | 109842329 | G | A |
| 11 | 109843953 | A | G |
| 11 | 109869302 | A | G |
| 11 | 109883416 | A | G |
| 11 | 109888898 | G | A |
| 11 | 109906198 | G | A |
| 11 | 109917159 | G | A |
| 11 | 109928038 | G | A |
| 11 | 109943972 | G | A |
| 11 | 109947509 | G | A |
| 11 | 109956346 | A | G |
| 11 | 109995642 | C | A |
| 11 | 109995944 | G | A |
| 11 | 109999257 | A | G |
| 11 | 110005581 | G | A |
| 11 | 110008339 | G | A |
| 11 | 110023164 | C | A |
| 11 | 110036365 | C | A |
| 11 | 110042157 | A | G |
| 11 | 110052890 | G | A |
| 11 | 110059838 | G | A |
| 11 | 110097651 | G | A |
| 11 | 110122426 | A | G |
| 11 | 110147873 | G | A |
| 11 | 110166131 | A | G |
| 11 | 113137365 | G | A |
| 12 | 54357198 | C | A |
| 12 | 54367690 | A | G |
| 12 | 56699623 | A | G |
| 12 | 56755058 | G | A |
| 12 | 68388126 | A | G |
| 12 | 68390382 | G | A |
| 12 | 68409009 | A | G |
| 12 | 68410080 | A | G |
| 12 | 68421622 | A | G |

| | | | |
|---|---|---|---|
| 12 | 68434014 | A | G |
| 12 | 68447643 | A | G |
| 12 | 68457881 | G | A |
| 12 | 68458643 | A | G |
| 12 | 68461640 | A | G |
| 12 | 68464082 | A | G |
| 12 | 68474095 | A | G |
| 12 | 68485984 | G | A |
| 12 | 68486435 | C | A |
| 12 | 68490340 | G | A |
| 12 | 68495027 | A | G |
| 12 | 68496077 | A | G |
| 12 | 68500075 | A | G |
| 12 | 68504592 | A | G |
| 12 | 68564677 | G | A |
| 12 | 68565441 | G | A |
| 12 | 68572933 | A | G |
| 12 | 68582172 | G | A |
| 12 | 68591685 | G | A |
| 12 | 68603834 | A | G |
| 12 | 68616795 | G | A |
| 12 | 68617219 | G | A |
| 12 | 120915220 | A | C |
| 12 | 120933977 | G | A |
| 12 | 125511207 | G | A |
| 13 | 21348682 | A | G |
| 14 | 101248696 | A | C |
| 14 | 101251989 | A | G |
| 14 | 101268562 | A | G |
| 14 | 101275683 | G | A |
| 14 | 101275950 | A | G |
| 14 | 101297515 | G | A |
| 14 | 101301012 | G | A |
| 14 | 101301866 | G | A |
| 14 | 101309759 | A | G |
| 14 | 101312869 | A | G |
| 14 | 101313093 | A | C |
| 14 | 101323211 | A | G |
| 16 | 11154770 | A | G |
| 16 | 28937275 | A | G |
| 16 | 31364909 | A | G |
| 16 | 31394722 | A | G |
| 16 | 31407528 | A | G |
| 16 | 79692668 | G | A |
| 16 | 79699967 | G | A |

| | | | |
|---|---|---|---|
| 16 | 79704152 | A | G |
| 16 | 79709503 | A | G |
| 16 | 79728572 | G | A |
| 16 | 79729031 | G | A |
| 16 | 79738425 | G | A |
| 16 | 79738625 | A | G |
| 16 | 79740249 | G | A |
| 16 | 79747885 | G | A |
| 16 | 79749276 | A | G |
| 16 | 79749341 | A | G |
| 16 | 79749353 | G | A |
| 16 | 79750512 | G | A |
| 16 | 79777617 | A | G |
| 16 | 79779098 | A | G |
| 16 | 79794709 | A | G |
| 16 | 79798321 | C | A |
| 17 | 58168322 | A | G |
| 18 | 52599038 | A | G |
| 18 | 52604083 | G | A |
| 18 | 53125364 | C | A |
| 18 | 53167520 | A | G |
| 18 | 53434122 | G | A |
| 18 | 53454369 | A | C |
| 18 | 53742508 | A | G |
| 18 | 53755311 | A | G |
| 18 | 54265847 | A | G |
| 19 | 6735929 | A | G |
| 19 | 6736607 | G | A |
| 19 | 47597102 | A | G |
| 19 | 47615835 | A | G |
| 19 | 47627065 | G | A |
| 19 | 47636827 | C | A |
| 19 | 47642780 | A | G |
| 19 | 47658320 | G | A |
| 19 | 47661493 | A | G |
| 19 | 47700860 | G | A |
| 19 | 47740491 | A | G |
| 19 | 47742310 | A | G |
| 19 | 47767643 | A | G |
| 19 | 47774706 | A | G |
| 19 | 47774772 | G | A |
| 19 | 47787587 | A | G |
| 19 | 47793185 | A | G |
| 19 | 47795969 | A | G |
| 19 | 47812900 | G | A |

| | | | |
|---|---|---|---|
| 19 | 47823871 | C | A |
| 19 | 55216162 | A | G |
| 19 | 55216894 | A | G |
| 19 | 55216950 | A | G |
| 19 | 55228948 | A | G |
| 19 | 55237616 | A | G |
| 19 | 55248107 | A | C |
| 20 | 34618622 | C | G |
| 20 | 34647457 | G | A |
| 20 | 44834924 | G | A |
| 20 | 44996450 | G | A |
| 20 | 45014787 | A | C |
| 20 | 56273207 | A | G |
| 20 | 56289402 | A | C |
| 21 | 46341197 | G | A |
| 21 | 46344426 | A | G |
| 21 | 46349496 | A | G |

**Supp. Table 3. Immune-related lncRNA SNPs in significant associations with endemic PF at the additive model.**

| SNP | Genomic position (hg19) | Odds Ratio | L95 | U95 | *p* value | lncRNA genes | Overlapped coding genes |
|---|---|---|---|---|---|---|---|
| rs3806156*A* | Chr6:32373698 | 0.6075 | 0.4506 | 0.8192 | 0.001083 | *lnc-BTNL2-1 (TSBP1-AS1)* | *BTNL2* |
| rs11100883*A* | Chr4:146450970 | 1.657 | 1.215 | 2.26 | 0.00142 | *lnc-SMAD1* | *SMAD1* |
| rs28362682*T* | Chr6:32372863 | 0.4627 | 0.2853 | 0.7502 | 0.001778 | *lnc-BTNL2-1 (TSBP1-AS1)* | *BTNL2* |
| rs11539588*A* | Chr19:6735929 | 0.3849 | 0.189 | 0.7837 | 0.008499 | *lnc-C3-2* | *GPR108* |
| rs2473122*A* | Chr6:22312287 | 1.721 | 1.138 | 2.601 | 0.01006 | *CASC15* | |
| rs1317186 *A* | Chr8:53318644 | 1.487 | 1.094 | 2.021 | 0.0112 | *lnc-ST18* | *ST18* |
| rs10228040*A* | Chr7:130653616 | 1.465 | 1.089 | 1.971 | 0.01155 | *LINC-PINT* | |
| rs9643471*A* | Chr8:53259979 | 0.5507 | 0.3466 | 0.8751 | 0.01158 | *lnc-ST18* | *ST18* |
| rs7812207*A* | Chr7:130784051 | 0.6139 | 0.418 | 0.9016 | 0.01283 | *LINC-PINT* | |
| rs12192707*A* | Chr6:6479161 | 0.68 | 0.5007 | 0.9237 | 0.01359 | *LY86-AS1* | |
| rs7386237*G* | Chr8:53262036 | 0.6082 | 0.4045 | 0.9143 | 0.01682 | *lnc-ST18* | *ST18* |
| rs4455715*A* | Chr6:6410663 | 1.425 | 1.066 | 1.906 | 0.01686 | *LY86-AS1* | |
| rs10404456*G* | Chr19:47812900 | 1.425 | 1.065 | 1.908 | 0.01721 | *lnc-C5AR1* | *C5AR1* |
| rs742286*A* | Chr6:22391406 | 1.474 | 1.067 | 2.037 | 0.01861 | *CASC15* | |
| rs10259462*G* | Chr7:130706995 | 1.518 | 1.071 | 2.154 | 0.01915 | *LINC-PINT* | |
| rs1418358*A* | Chr9:29132164 | 1.418 | 1.056 | 1.905 | 0.02025 | *lnc-LINGO2-5* | |
| rs13179150*A* | Chr5:39340135 | 1.464 | 1.059 | 2.023 | 0.0212 | *lnc-C9-1* | *C9* |
| rs7847094*G* | Chr9:29119535 | 0.7126 | 0.5315 | 0.9555 | 0.02358 | *lnc-LINGO2-5* | |
| rs2294461*C* | Chr6:6614501 | 1.669 | 1.071 | 2.601 | 0.02373 | *LY86-AS1* | *LY86* |
| rs674485*G* | Chr11:65197393 | 1.381 | 1.044 | 1.827 | 0.02381 | *NEAT1* | |
| rs7468614*A* | Chr9:29054853 | 1.428 | 1.048 | 1.945 | 0.02398 | *lnc-LINGO2-5* | |
| rs1294468 *G* | Chr6:6544702 | 1.375 | 1.042 | 1.816 | 0.02451 | *LY86-AS1* | |
| rs7932174*A* | Chr11:110122426 | 0.6949 | 0.5058 | 0.9546 | 0.02465 | *lnc-ZC3H12C* | *RDX* |
| rs3763308*A* | Chr6:32374640 | 0.5315 | 0.3024 | 0.9344 | 0.02812 | *TSBP-AS1* | *BTNL2* |
| rs4712624*G* | Chr6:21693152 | 1.442 | 1.038 | 2.003 | 0.02887 | *CASC15* | |
| rs1884537*A* | Chr14:101251989 | 0.7355 | 0.5582 | 0.969 | 0.029 | *MEG3* | |
| rs7767991*A* | Chr6:22380084 | 0.7145 | 0.5271 | 0.9684 | 0.03026 | *CASC15* | |
| rs13242887*G* | Chr7:130751832 | 0.6786 | 0.4753 | 0.9688 | 0.03282 | *LINC-PINT* | |
| rs12678991*A* | Chr8:53320710 | 1.351 | 1.025 | 1.78 | 0.03299 | *lnc-ST18-4* | *ST18* |
| rs3024493*A* | Chr1:206943968 | 0.5874 | 0.3584 | 0.9625 | 0.03473 | *lnc-IL10-5* | *IL10* |
| rs498688*G* | Chr3:187679835 | 1.348 | 1.02 | 1.782 | 0.03569 | *LINC01991* | |
| rs11944685*A* | Chr4:146414267 | 0.7402 | 0.5573 | 0.983 | 0.03766 | *lnc-SMAD1* | *SMAD1* |
| rs7250152*A* | Chr19:47795969 | 1.391 | 1.018 | 1.899 | 0.03805 | *lnc-C5AR1* | *C5AR1* |
| rs12376696*A* | Chr9:29024733 | 1.425 | 1.018 | 1.994 | 0.0388 | *lnc-LINGO2-5* | |
| rs7768473*A* | Chr6:22431621 | 0.7327 | 0.5446 | 0.9858 | 0.03992 | *CASC15* | |
| rs7870900*G* | Chr9:29153469 | 1.359 | 1.014 | 1.822 | 0.04035 | *lnc-LINGO2-5* | |
| rs2870957*C* | Chr12:68486435 | 0.713 | 0.5157 | 0.9859 | 0.04077 | *IFNG-AS1* | |
| rs6974804*G* | Chr7:130783121 | 1.367 | 1.011 | 1.848 | 0.04232 | *LINC-PINT* | |
| rs1047581*G* | Chr11:33724992 | 0.7233 | 0.5282 | 0.9904 | 0.04338 | *AL049629.2* | *CD59* |
| rs10853784*A* | Chr19:47793185 | 0.7494 | 0.5661 | 0.992 | 0.04382 | *lnc-C5AR1* | |

| rs7742855*A | Chr6:22044638 | 1.354 | 1.007 | 1.82 | 0.04477 | *CASC15* | |
|---|---|---|---|---|---|---|---|
| rs217727*A | Chr11:2016908 | 0.6995 | 0.493 | 0.9924 | 0.0452 | *H19* | |
| rs12659504*G | Chr5:148805005 | 1.533 | 1.008 | 2.333 | 0.04603 | *lnc-IL17B-5* | |
| rs17142362*A | Chr6:6601050 | 0.5542 | 0.3097 | 0.9917 | 0.04681 | *LY86-AS1* | *LY86* |
| rs700218*A | Chr5:39354542 | 0.1195 | 0.01469 | 0.9727 | 0.04705 | *lnc-C9* | *C9* |
| rs2918284*A | Chr5:148588846 | 0.7498 | 0.5641 | 0.9966 | 0.04731 | *lnc-IL17B-3* | *ABLIM3* |
| rs1418344*G | Chr9:29084708 | 1.342 | 1.002 | 1.797 | 0.04851 | *lnc-LINGO2-5* | |

Associated SNPs: $p < 0.05$. L95 - lower limitit of 95% condifence interval; U95 - upper limit of 95% confidence interval.

**Supp. Table 4. Computational annotations and in silico predictions for the top-4 PF associated SNPs.**

| Associated SNP | | The associated SNP impacts on | | Linkage Disequilibrium | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **SNP impact** | **Scores** | **LD SNPs** | **Position (GRCh37)** | **MAF** | **D'** | **r²** | **Correlated Alleles** | **RegulomeDB score** |
| **rs2473122 (CASC15)** | **Alleles** | Alters TF binding motifs | **RegulomeDB** | rs2504809 | 6:22312399 | 7,2% | 1.0 | **1.0** | A=A,G=C | 6 |
| | G>A>T | | 6 | rs2655433 | 6:22312262 | 7,2% | 1.0 | **1.0** | A=T,G=C | 7 |
| | **MAF** | | **CADD** | rs2655416 | 6:22321054 | 8,6% | 1.0 | **0.8204** | A=A,G=G | 2b |
| | Total: 21% (A) | **Gene splicing prediction** | G:2.560; T:2.131 | rs715226 | 6:22317563 | 8,6% | 1.0 | **0.8204** | A=C,G=T | 6 |
| | EUR: 7,3% | Creation of an intronic ESE site | **GERP** | rs2744105 | 6:22317319 | 8,6% | 1.0 | **0.8204** | A=C,G=G | 7 |
| | AFR: 29% | | 1,29 | | | | | | | |
| **rs10228040 (LINC-PINT)** | **Alleles** | - | **RegulomeDB** | rs2048672 | 7:130653851 | 28,8% | 1.0 | **1.0** | G=C,A=A | 7 |
| | G>A | | 4 | rs13225636 | 7:130654398 | 28,8% | 1.0 | **1.0** | G=G,A=A | 5 |
| | **MAF** | | **CADD** | | | | | | | |
| | Total: 51% (A) | **Gene splicing prediction** | A:0.380 | | | | | | | |
| | EUR: 28,6% | - | **GERP** | | | | | | | |
| | AFR: 81% | | -3.79 | | | | | | | |
| **rs7812207 (LINC-PINT)** | **Alleles** | - | **RegulomeDB** | No SNPs found with $r^2$>0.8 in selected populations | | | | | | |
| | G>A>T | | - | | | | | | | |
| | **MAF** | | **CADD** | | | | | | | |
| | Total: 23% (T) | **Gene splicing prediction** | A:0.481, T:0.367 | | | | | | | |
| | EUR: 19,6% | EIE site | **GERP** | | | | | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | AFR: 16% | broken | -1.58 | | | | | | | |
| | **Alleles** | Present in regulatory region active on CD4+ T cells | **RegulomeDB** | rs12205172 | 6:6480002 | 40,3% | 1.0 | **1.0** | G=A,A=C | 4 |
| | G>A | | 6 | rs12215071 | 6:6485608 | 40,7% | 0.9732 | **0.9285** | G=A,A=G | 5 |
| **rs12192707 (LY86-AS1)** | **MAF** | | **CADD** | rs7740245 | 6:6487246 | 40,7% | 0.9732 | **0.9285** | G=G,A=A | 7 |
| | Total: 22% (A) | Present in regulatory region active on CD4+ T cells | A: 0.499 | rs6916258 | 6:6490534 | 40,7% | 0.9665 | **0.9158** | G=T,A=C | 6 |
| | EUR: 40,3% | | **GERP** | rs6901499 | 6:6477441 | 40,3% | 0.9535 | **0.9092** | G=T,A=C | 7 |
| | AFR: 13% | | -4.84 | rs12193827 | 6:6488577 | 39,9% | 0.9531 | **0.8965** | G=A,A=G | 5 |
| | | **Gene splicing prediction** | | rs12203144 | 6:6477912 | 42,3% | 0.9794 | **0.8804** | G=C,A=A | 7 |
| | | Creation of EIE and ESE sites; ESS site broken | | rs11755432 | 6:6468046 | 38,5% | 0.9444 | **0.8287** | G=A,A=T | 5 |
| | | | | rs4959384 | 6:6467773 | 38,5% | 0.9444 | **0.8287** | G=C,A=T | 3a |
| | | | | rs7748193 | 6:6466847 | 38,5% | 0.9444 | **0.8287** | G=T,A=A | 7 |
| | | | | rs7760763 | 6:6462911 | 38,5% | 0.9444 | **0.8287** | G=A,A=G | 7 |
| | | | | rs9504825 | 6:6481568 | 44,6% | 0.9928 | **0.8261** | G=A,A=G | 6 |
| | | | | rs12190373 | 6:6465900 | 38,3% | 0.9442 | **0.8227** | G=C,A=T | 5 |

Computational predictions and annotations for the 4 most PF-associated SNPs, and their Linkage disequilibrium (LD) SNPs, considering $r^2 > 0.8$. Underlined alleles are the ancestral alleles. The evaluated European population (EUR) is composed by CEU (Utah residents with European ancestry), TSI (Toscani) and IBS (Iberian), while YRI (Yoruba) was considered for the African population (AFR). MAF - minor allele frequency; CADD - Combined Annotation Dependent Depletion score; GERP - Genomic Evolutionary Rate Profiling score; ESE – exonic splicing enhancers; ESS - exonic splicing silencers; EIE - exon-identity elements; TF – transcription factors. Sources: Ensembl, Regulome DB, Human Splicing Finder, lncRNASNP2 and LDlink online tools.

Supp Table 5.xlsx

**Supp. Table 6.** Primer sequences used for iPlex genotyping.

| SNP | Forward Primer Sequence | Reverse Primer Sequence | Extended Primer Sequence |
|---|---|---|---|
| *rs10228040* | ACGTTGGATGTCATTCTCAGGTGTTGGTTC | ACGTTGGATGACAGGATTCACTGGTTTGGG | caTGGAAGATGTGTTAGAAATGAC |
| *rs12192707* | ACGTTGGATGAAGCAGGAGCTCAAGACACT | ACGTTGGATGCAACCCCATCAACTAATGTG | GGTTTATGTCTGCCTCTCT |
| *rs674485* | ACGTTGGATGGGCATGGTGCTCTCAGAAC | ACGTTGGATGCTTCTGAATTGAACCCTGCC | aCCTGTGAGGGTGGTT |
| *rs1884537* | ACGTTGGATGTGGAGCAGACTAAAATGAGC | ACGTTGGATGTGCTCAGTCACCACACAGTC | ccccACTAAAATGAGCCAACAGG |

## 7. CONCLUSÃO FINAL

A partir dos resultados obtidos nos trabalhos desta tese, concluímos que genes de lncRNAs podem estar relacionados com diferentes processos imunológicos e autoimunes do pênfigo. Foram demonstrados que lncRNAs imuno-relacionados estão associados com o PF: seja para ambas as formas (endêmica e esporádica) da doença, como é o caso do lncRNA *LINC-PINT*; ou para alguma forma específica de PF, como é o caso da associação de *LY86-AS1* em PFE.

Já os lncRNAs da família dos HCGs, apesar de ainda pouco conhecidos, estão geneticamente associados com os pênfigos foliáceo e vulgar. Destes, destacamos os genes *TSBP1-AS1* (que abriga o gene *HCG23*) e *HCG27*, associados com ambas as doenças. Além disso, *TSBP1-AS1* e *HCG27* estão diferencialmente expressos em PV e PF/PFE, respectivamente. Logo, sugerimos ambos como fortes genes candidatos para posteriores estudos em pênfigo ou em outras doenças autoimune semelhantes, a fim de validar as associações encontradas em diferentes coortes e explorar seus mecanismos moleculares, ampliando a compreensão sobre os aspectos funcionais e regulatórios destas moléculas, nos contextos fisiológicos e imunopatológicos.

## 8. REFERÊNCIAS


AMAGAI, M. et al. The clinical phenotype of pemphigus is defined by the anti-desmoglein autoantibody profile. **Journal of the American Academy of Dermatology**, v. 40 (2), p. 167–170, 1999.

AMORMINO, S. A. F.; BARBOSA, A. A. M. Pênfigo vulgar: revisão de literatura e relato de caso clínico. **Revista Periodontia**, v. 20, n° 02, 2010.

ANAPARTI, V. et al. Whole Blood Targeted Bisulfite Sequencing Validates Differential Methylation in C6ORF10 gene of Patients with Rheumatoid Arthritis. **The Journal of Rheumatology**, November 2019, jrheum.190376; DOI: https://doi.org/10.3899/jrheum.190376

AOKI, V. et al. Endemic pemphigus foliaceus (fogo selvagem) and pemphigus vulgaris: immunoglobulin G heterogeneity detected by indirect immunofluorescence. **Revista do Hospital das Clínicas**, 59(5), 251-256, 2004. https://dx.doi.org/10.1590/S0041-87812004000500005.

AOKI, V.; SOUSA, J. X.; DIAZ, L. A. Pathogenesis of Endemic Pemphigus Foliaceus. **Dermatologic Clinics**, V. 29 (3), p. 413–418, 2011.

AUGUSTO, D.G. et al. Activating KIR and HLA Bw4 ligands are associated to decreased susceptibility to pemphigus foliaceus, an autoimmune blistering skin disease. **PLoS One**, v. 7 (7), 2012.

BIRD, C.E. et al. Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. **PLoS Genet.**, v. 6, 2010.

BRAUN-PRADO, K.; PETZL-ERLER, M. L. Programmed cell death 1 gene (PDCD1) polymorphism and pemphigus foliaceus (fogo selvagem) disease susceptibility. **Genetics and Molecular Biology**, v. 30, p. 314–321, 2007.

BROCHADO, M.J. et al. Differential HLA class I and class II associations in pemphigus foliaceus and pemphigus vulgaris patients from a prevalent Southeastern Brazilian region. **J Autoimmun.**, Aug;72:19-24, 2016. doi: 10.1016/j.jaut.2016.04.007.

BUMILLER-BINI, V. et al. Sparking Fire Under the Skin? Answers From the Association of Complement Genes With Pemphigus Foliaceus. **Frontiers in immunology**, 9, 695, 2018. doi:10.3389/fimmu.2018.00695.

BUMILLER-BINI, V. et al. Condemned or Not to Die? Gene Polymorphisms Associated With Cell Death in Pemphigus Foliaceus. **Front. Immunol.,** 18 October 2019, https://doi.org/10.3389/fimmu.2019.02416

BYSTRYN, J. C.; RUDOLPH, J. L. Pemphigus.**The Lancet**, v. 366 (9479), p. 61–73, 2005.

CAMARGO, C. M.; AUGUSTO, D. G.; PETZL-ERLER, M. L. Differential gene expression levels might explain association of LAIR2 polymorphisms with pemphigus. **Human Genetics**, v. 135 (2), p. 233-244, 2016.

CAMPBELL, I. et al. Pênfigo foliáceo endêmico: fogo selvagem / endemic pemphigusfoliaceous: fogo selvagem. **Anais brasileiros de dermatologia**, v. 76(1), p. 13-33, 2001.

CAPRONI, M. et al. Further support for a role for Th2-like cytokines in blister formation of pemphigus. **Clin Immunol.**, Feb;98(2):264-71, 2001.

CARNINCI, P. et al. The transcriptional landscape of the mammalian genome. **Science**, v. 309, 1559–1563, 2005.

CHEN, Y. et al. Long non-coding RNA HCG11 modulates glioma progression through cooperating with miR-496/CPEB3 axis. **Cell Prolif.**, Sep;52(5):e12615. 2019. doi: 10.1111/cpr.12615.

CHUNG, S.A. et al. Lupus nephritis susceptibility loci in women with systemic lupus erythematosus. **Journal of the American Society of Nephrology**: JASN, 25(12), 2859–2870, 2014. doi:10.1681/ASN.2013050446

CIPOLLA, G.A. et al. A 3′UTR polymorphism marks differential KLRG1 mRNA levels through disruption of a miR-584-5p binding site and associates with pemphigus foliaceus susceptibility. **Biochimica et Biophysica Acta**, v. 1859, p. 1306-1313, 2016.

CIPOLLA G. A.; PETZL-ERLER, M. L. **Análise de expressão de microRNAs candidatos à regulação da resposta imune adaptativa em pênfigo foliáceo**. Dissertação de mestrado. Universidade Federal do Paraná, Curitiba, 2012.

CIPOLLA, G.A. et al. Long Non-Coding RNAs in Multifactorial Diseases: Another Layer of Complexity. **Non-coding RNA**, 4(2), 13, 2018.

CLANCY, R.M. et al. Identification of candidate loci at 6p21 and 21q22 in a genome-wide association study of cardiac manifestations of neonatal lupus. **Arthritis and rheumatism**, 62(11), 3415–3424, 2010. doi:10.1002/art.27658

CLOP, A. et al. An in-depth characterization of the major psoriasis susceptibility locus identifies candidate susceptibility alleles within an HLA-C enhancer element. **PloS one**, 8(8), e71690, 2013. doi:10.1371/journal.pone.0071690

CRICK, F.H. et al. General nature of the genetic code for proteins. **Nature**, v. 192, p. 1227–1232, 1961.

CULTON, D.A. et al. Advances in Pemphigus and its Endemic Pemphigus Foliaceus (FogoSelvagem) Phenotype: A Paradigm of Human Autoimmunity. **Journal of Autoimmunity**, v. 31 (4), p. 311–324, 2008.

DALLA-COSTA R. et al. Polymorphisms in the 2q33 and 3q21 chromosome regions including Tcell coreceptor and ligand genes may influence susceptibility to pemphigus foliaceus. **Human Immunology**, v. 71, p. 809, 2010.

DE SENA, L.N.M. et al. HLA class II alleles of susceptibility and protection in Brazilian and Dutch pemphigus foliaceus. **Br J Dermatol.,** 178: e212–14, 2018.

DEBIEC, H. et al. Transethnic, Genome-Wide Analysis Reveals Immune-Related Risk Alleles and Phenotypic Correlates in Pediatric Steroid-Sensitive Nephrotic Syndrome. **Journal of the American Society of Nephrology** : JASN, 29(7), 2000–2013, 2018. doi:10.1681/ASN.2017111185

DERRIEN, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. **Genome Res**., v. 22, p.1775–1789, 2012.

DEYHIMI, P.; ALISHAHI, B. Study of extrinsic apoptotic pathway in oral pemphigus vulgaris using TNFR 1 and FasL immunohistochemical markers and TUNEL technique. **J Dent.** 19:132–41, 2018.

DIAZ, L.A. et al. Endemic pemphigus foliaceus (fogo selvagem). I. Clinical features and immunopathology. **Journal of the American Academy of Dermatology**, V. 20 (4), P. 657–669, 1989.

DIAZ, L.A. et al. [2]. Endemic Pemphigus Foliaceus (Fogo Selvagem): II. Current and Historic Epidemiologic Studies. **Journal of Investigative Dermatology**, v. 92, p. 4–12, 1989.

DIAZ, L.A. et al. The IgM anti-desmoglein 1 response distinguishes Brazilian pemphigus foliaceus (fogo selvagem) from other forms of pemphigus. **J Invest Dermatol.,** 128:667-75, 2008.

DINGER, M.E. et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. **Genome Res**., v. 18, p. 1433–1445, 2008.

EBERHARD, Y. et al. Cytokine polymorphisms in patients with pemphigus. **Arch Dermatol Res.**, Jan;296(7):309-13, 2005.

ESTELLER, M. Non-coding RNAs in human disease. **Nature Reviews Genetics**, v. 12, p. 861-874, 2011.

FANG, S. et al. NONCODEV5: a comprehensive annotation database for long non-coding RNAs. **Nucleic Acids Res**, 2017.

FARIAS, T.D.J.; AUGUSTO, D.G.; DE ALMEIDA, R.C.; MALHEIROS, D.; PETZL-ERLER, M.L. Screening the full leucocyte receptor complex genomic region revealed associations with pemphigus that might be explained by gene regulation. **Immunology**, 156(1):86-93, 2019.

FELICIANI, C. et al. In vitro and in vivo expression of interleukin-1alpha and tumor necrosis factor-alpha mRNA in pemphigus vulgaris: interleukin-1alpha and tumor necrosis factor-alpha are involved in acantholysis. **J Invest Dermatol.,** 114:71–7, 2000. 10.1046/j.1523-1747.2000.00835.x

FENG, B. J. et al. Multiple Loci within the major histocompatibility complex confer risk of psoriasis. **PLoS genetics**, 5(8), e1000606, 2009. doi:10.1371/journal.pgen.1000606

FRUSIC-ZLOTKIN, M. et al. The interaction of pemphigus autoimmunoglobulins with epidermal cells. **Ann NY Acad Sci**, v. 1050, p. 371-79, 2005.

GENG, H.; TAN, X.D. Functional diversity of long non-coding RNAs in immune regulation. **Genes Dis.**, v. 3(1), p. 72–81, 2016.

GIBB, E.A.; BROWN, C.J.; LAM, W.L. The functional role of long non-coding RNA in human carcinomas. **Mol. Cancer**, v. 10, p. 38, 2011.

GOLDBERG, A.C.; RIZZO, L.V. Estrutura do MHC e função – apresentação de antígenos. **Einstein.**, 13(1):153-6, 2015.

GRANDO S.A. Pemphigus autoimmunity: hypotheses and realities. **Autoimmunity**, 45(1), 7–35, 2012. https://doi.org/10.3109/08916934.2011.606444

GUTTMAN, M. et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. **Nature**, v. 458, p. 223–227, 2009.

HANS-FILHO, G. et al. The cooperative group on fogo selvagem research endemic pemphigus foliaceous (fogo selvagem). **Clinical Dermatology**, v. 17, p. 225-235, 1999.

HARROW, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. **Genome Research**. v. 22 p.1760–1774, 2012.

HERTL, M.; EMING, R.; VELDMAN, C. T cell control in autoimmune bullous skin disorders. **The Journal of Clinical Investigation**, v. 116 (5), p. 1159-1166, 2006.

HOLM, S.J. et al. HLA-Cw*0602 associates more strongly to psoriasis in the Swedish population than variants of the novel 6p21.3 gene PSORS1C3. **Acta Derm. Venereol.**, 85, 2–8, 2005.

HORTON, R. et al. Gene map of the extended human MHC**. Nat Rev Genet.** Dec;5(12):889-99, 2004.

HOWE, K.L. et al. Ensembl Genomes 2020-enabling non-vertebrate genomic research. **Nucleic Acids Research** 2019. https://doi.org/10.1093/nar/gkz890

HRDLICKOVA, B. et al. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. **Biochimica et Biophysica Acta**, v.1842, p.1910-1922, 2014.

JOHNSSON, P. et al. “Evolutionary conservation of long non-coding RNAs; sequence, structure, function.” **Biochimica et biophysica acta**, vol. 1840,3: 1063-71, 2014. doi:10.1016/j.bbagen.2013.10.035

JOLY, P.; LITROWSKI, N. Pemphigus group (vulgaris, vegetans, foliaceus, herpetiformis, brasiliensis). **Clin Dermatol.**, 29(4):432-6, 2001.

JORDON, R.E. et al. Classical and alternate pathway activation of complement in pemphigus vulgaris lesions. **J Investig Dermatol.**, 63:256–9, 1974. 10.1111/1523-1747.ep12680098

KAHLOUN, A.E. et al. Localization of seven new genes around the HLA-A locus. **Human Molecular Genetics**, 2(1), 55–60, 1993. doi:10.1093/hmg/2.1.55

KAPRANOV, P. et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. **Science**, v. 316, p.1484–1488, 2007.

KASPERKIEWICZ, M. et al. Pemphigus. **Nat Rev Dis Prim**, 3(1):17026, 2017. Available from: http://www.nature.com/articles/nrdp201726

KAWANA, S. et al. Complement fixation by Brazilian Pemphigus foliaceus autoantibodies. **Clinical and Experimental Immunology**, v. 71(3), p. 464–469, 1988.

KAWANA, S. et al. Deposition of membrane attack complex of complement in pemphigus vulgaris and pemphigus foliaceous skin. **The Society of Investigative Dermatology**, v. 92, p. 588-591, 1989.

KHONG, J.J. et al. Pooled genome wide association detects association upstream of FCRL3 with Graves' disease. **BMC genomics**, 17(1), 939, 2016. doi:10.1186/s12864-016-3276-z.

KINO, T. et al. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. **Sci. Signal.**, v. 3, p. ra8, 2010.

KITAJIMA, Y.; AOYAMA, Y. A Perspective of Pemphigus from Bedside and Laboratory-Bench. **Clinical Reviews in Allergy & Immunology**, v. 33 (1), p. 57-66, 2007.

KNIGHT, J. et al. Conditional analysis identifies three novel major histocompatibility complex loci associated with psoriasis. **Human molecular genetics**, 21(23), 5185–5192, 2012. doi:10.1093/hmg/dds344

KORMAN, N. Pemphigus. **Journal of the American Academy of Dermatology**, v. 18 (6), p. 1219–1238, 1988.

KRIDIN K. Pemphigus group: overview, epidemiology, mortality, and comorbidities. **Immunol Res**; 66: 255–70, 2018.

KULSKI, J.K. Long Noncoding RNA HCP5, a Hybrid HLA Class I Endogenous Retroviral Gene: Structure, Expression, and Disease Associations. **Cells**, 8(5) 480, 2019. https://doi.org/10.3390/cells8050480

KULSKI, J.K.; SHIINA, T.; DIJKSTRA, J.M. Genomic Diversity of the Major Histocompatibility Complex in Health and Disease. **Cells**, 8(10) 1270, 2019. https://doi.org/10.3390/cells8101270

KUMAR, V.; WIJMENGA, C. From genome-wide association studies to disease mechanisms: celiac disease as a model for autoimmune diseases. **Semin Immunopathol.**, v. 34(4), p. 567–580, 2012.

LEE, H.E. et al. Biphasic activation of p38MAPK suggests that apoptosis is a downstream event in pemphigus acantholysis. **J Biol Chem.**, 284:12524–32, 2009. doi: 10.1074/jbc.M808204200

LI, S. et al. Association between HLA-DQB1 polymorphisms and pemphigus vulgaris: A meta-analysis**. Immunol Invest.**, Jan;47(1):101-112, 2018. doi: 10.1080/08820139.2017.1385622.

Lin, X. et al. Susceptibility Genes for Multiple Sclerosis Identified in a Gene-Based Genome-Wide Association Study. **Journal of clinical neurology** (Seoul, Korea), 11(4), 311–318, 2015. doi:10.3988/jcn.2015.11.4.311

LOBO-ALVES, S.C. et al. Long non-coding RNA polymorphisms influence susceptibility to endemic pemphigus foliaceus. **Br J Dermatol.**, 2019.

LOBO-ALVES, S. C. et al. [2]. Region 1p13.2 including the RSBN1, PTPN22, AP4B1 and long non-coding RNA genes does not bear risk factors for endemic pemphigus foliaceus (fogo selvagem). **International Journal of Immunogenetics**, 2019. doi:10.1111/iji.12423

LUO, M. et al. IFNA-AS1 regulates CD4+ T cell activation in myasthenia gravis though HLA-DRB1. **Clin Immunol.,** Oct;183:121-131, 2017. doi: 10.1016/j.clim.2017.08.008.

LOMBARDI, M.L. et al. Common human leukocyte antigen alleles in pemphigus vulgaris and pemphigus foliaceus Italian patients. **J Invest Dermatol**. Jul;113(1):107-10, 1999.

MAHONEY, M. G. et al. Explanations for the clinical and microscopic localization of lesions in pemphigus foliaceus and vulgaris. **The Journal of Clinical Investigation**, v. 103 (4), p. 461-468, 1999.

MAHONEY, M.G.; WANG, Z.H.; STANLEY, J.R. Pemphigus vulgaris and pemphigus foliaceus antibodies are pathogenic in plasminogen activator knockout mice. **The Journal of Investigative Dermatology**, v. 113, n. 1, p. 22-5, 1999.

MALAVIA, T.A. et al. Generating testable hypotheses for schizophrenia and rheumatoid arthritis pathogenesis by integrating epidemiological, genomic, and protein interaction data. **NPJ schizophrenia**, 3, 11, 2017. doi:10.1038/s41537-017-0010-z

MALHEIROS, F. D.; PETZL-ERLER, M. L. Individual and epistatic effects of genetic polymorphisms of B-cell co-stimulatory molecules on susceptibility to pemphigus foliaceus. **Genes and Immunity**, v. 10, p. 547–558, 2009.

MALHEIROS. F., D. et al. Genome-wide gene expression profiling reveals unsuspected molecular alterations in pemphigus foliaceus. **Immunology**, 2014.

NIENS, M. et al. The Human Leukocyte Antigen Class I Region Is Associated with EBV-Positive Hodgkin's Lymphoma: HLA-A and HLA Complex Group 9 Are Putative Candidate Genes. **Cancer Epidemiol Biomarkers Prev**,15(11):2280–4, 2006.

MAURANO, M.T. et al. Systematic localization of common disease-associated variation in regulatory DNA. **Science** (New York, N.Y.), 337(6099), 1190–1195, 2012. doi:10.1126/science.1222794

MEDER, B. et al. A genome-wide association study identifies 6p21 as novel risk locus for dilated cardiomyopathy. **European Heart Journal**, 35(16), 1069–1077, 2013. doi:10.1093/eurheartj/eht251

MEYERA, N.; MISERY L. Geoepidemiologic considerations of auto-immune pemphigus. **Autoimmunity Reviews**, v. 9 (5), p. A379–A382, 2010.

MERCER, T.R. et al. Specific expression of long noncoding RNAs in the mouse brain. **Proc. Natl. Acad. Sci. U.S.A.**, v. 105, p. 716–721, 2008.

MESSIAS, I.T. et al. Complement activation in Brazilian pemphigus foliaceus. **Clinical and Experimental Dermatology**, v. 14, p. 51-55, 1989.

MESSIAS-REASON, I.J. et al. Circulating levels of mannan-binding lectin (MBL) and MBL-associated serine protease 2 in endemic pemphigus foliaceus. **Clinical and Experimental Dermatology**, v. 33, p. 495–497, 2008.

MESSIAS-REASON, I.J.; NISIHARA, R. M.; MOCELIN, V. Mannan-binding lectin and ficolin deposition in skin lesions of pemphigus. **Archives of Dermatological Research**, v. 303 (7), p. 521-525, 2011.

MILLIGAN, M.J.; LIPOVICH, L. Pseudogene-derived lncRNAs: emerging regulators of gene expression. **Frontiers in genetics**, 5, 476, 2015. doi:10.3389/fgene.2014.00476

MOON, C.M. et al. Deep Resequencing of Ulcerative Colitis-Associated Genes Identifies Novel Variants in Candidate Genes in the Korean Population. **Inflammatory Bowel Diseases**, 24(8), 1706–1717, 2018. doi:10.1093/ibd/izy122

MORAES, M.E. et al. An epitope in the third hypervariable region of the DRB1 gene is involved in the susceptibility to endemic pemphigus foliaceus (fogo selvagem) in three different Brazilian populations. **Tissue Antigens**. Jan;49(1):35–40, 1997.

MORAN, V.A.; PERERA, R.J.; KHALIL, A.M. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. **Nucleic Acids Research**, v.40, p.6391–6400, 2012.

MOSAAD, Y.M. et al. Tumor necrosis factor-alpha−308 G>A and interleukin-6−174 G>C promoter polymorphisms and pemphigus. **Hum Immunol.,** 73:560–5, 2012. 10.1016/j.humimm.2012.02.001

NAGEL, A. et al. Clinical activity of pemphigus vulgaris relates to IgE autoantibodies against desmoglein 3. **Clin Immunol.**, Mar; 134(3):320-30, 2010.

NISIHARA, R. M.; DE BEM, R. S.; HAUSBERGER, R.; ROXO, V. S.; PAVONI, D. P.; PETZL-ERLER, M. L.; MESSIAS-REASON, I. J. Prevalence of autoantibodies in patients with endemic pemphigus foliaceus (fogo selvagem). **Archives of Dermatological research**, v. 295, p. 133-137, 2003.

OLIVEIRA, L.C. et al. Complement Receptor 1 (CR1, CD35) Polymorphisms and Soluble CR1: A Proposed Anti-inflammatory Role to Quench the Fire of "Fogo Selvagem" Pemphigus Foliaceus. **Front. Immunol.**, 22 November 2019 **|** https://doi.org/10.3389/fimmu.2019.02585

OLIVEIRA-JR, J. V. **Perfil Clínico e imunológico dos pênfigos vulgar e foliáceo com envolvimento umbilical**. Dissertação de mestrado, Faculdade de Medicina, Universidade de São Paulo, São Paulo, 2012.

OLIVEIRA, J.C. et al. LncRNAs in Cancer: another layer of complexity. **Journal of Gene Medicine**, v.21(1), 2018.

PAVONI, D.P. et al. Dissecting the associations of endemic Pemphigus Foliaceus (FogoSelvagem) with HLA-DRB1 alleles and genotypes. **Genes and Immunity**, v. 4, p. 110–116, 2003.

PAVONI, D.P. et al. Polymorphism of the promoter region and exon 1 of the CTLA4 gene in endemic pemphigus foliaceus (fogo selvagem). **Braz J Med Biol Res.**, Sep;39(9):1227-32, 2006.

PAYNE, A.S.; STANLEY, J.R. Pemphigus. **Fitzpatrick's Dermatology in General Medicine**. 8th ed. Vol 1. p.586, 2012.

PEGUEROLES, C.; GABALDÓN, T. Secondary structure impacts patterns of selection in human lncRNAs. **BMC Biol**., 14, 60, 2016.

PEREIRA, N.F.; HANSEN, J. A.; LIN, M. T.; ROXO, V. M. M. S.; BRAUN, K.; PETZL-ERLER, M. L. Cytokine gene polymorphisms in endemic pemphigus foliaceus: a possible role for IL6 variants. **Cytokine**, v. 28, p. 233-241, 2004.

PETZL-ERLER, M.L.; SANTAMARIA, J. Are HLA class II genes controlling susceptibility and resistance to Brazilian pemphigus foliaceus (fogo selvagem)? **Tissue Antigens**, v. 33, p. 408–414, 1989.

PETZL-ERLER, M.L. Beyond the HLA polymorphism: A complex pattern of genetic susceptibility to pemphigus. **Genetics and Molecular Biology**, 2020 (aceito para publicação).

PICHON, L. et al. A new non-HLA multigene family associated with thePERB11 family within theMHC class I region. Immunogenetics, 44(4), 259–267, 1996. doi:10.1007/bf02602555

PICHON, L. et al. [2]. Structural Analysis of the HLA-A/HLA-F Subregion: Precise Localization of Two New Multigene Families Closely Associated with the HLA Class I Sequences. Genomics, 32(2), 236–244, 1996. doi:10.1006/geno.1996.0110

PIOVEZAN, B.Z.; PETZL-ERLER, M.L. Both Qualitative and quantitative genetic variation of MHC Class II molecules may influence susceptibility to autoimmune diseases: the case of endemic pemphigus foliaceus. **Human Immunology**, v.74, p.1134-1140, 2013.

POLLMANN, R. et al. Pemphigus: a Comprehensive Review on Pathogenesis, Clinical Presentation and Novel Therapeutic Approaches. Vol. 54, **Clinical Reviews in Allergy and Immunology**. Humana Press Inc.; 2018.

POLLOCK, R.A. et al. Epigenome-wide analysis of sperm cells identifies IL22 as a possible germ line risk locus for psoriatic arthritis. **PloS one**, 14(2), e0212043, 2019. doi:10.1371/journal.pone.0212043

PONJAVIC, J.; PONTING, C.P.; LUNTER, G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. **Genome Res.**, 17(5):556–565, 2007. doi:10.1101/gr.6036807

PONTING, C.P. et al. Evolution and Functions of Long Noncoding RNAs. **Cell**, v. 136(4), p. 629–641, 2009.

PUVIANI, M. et al. Fas ligand in pemphigus sera induces keratinocyte apoptosis through the activation of caspase-8. **J Invest Dermatol**. Jan;120(1):164-7, 2003.

QIAN, Y. et al. Cutting Edge: Brazilian Pemphigus Foliaceus Anti-Desmoglein 1 Autoantibodies Cross-React with Sand Fly Salivary LJM11 Antigen. **The Journal of Immunology**, v. 189 (4), p. 1535-1539, 2012.

QUINN, J.J. et al. “Rapid evolutionary turnover underlies conserved lncRNA-genome interactions.” **Genes & development**, vol. 30,2: 191-207, 2016. doi:10.1101/gad.272187.115

RICAÑO-PONCE, I.; WIJMENGA, C. Mapping of Immune-Mediated Disease Genes. **Annual Review of Genomics and Human Genetics**, v. 14, p. 325-353, 2013.

SAEIDI, L. et al. Long non-coding RNA LY86-AS1 and HCG27_201 expression in type 2 diabetes mellitus. **Molecular Biology Reports**, 2018. doi:10.1007/s11033-018-4429-8

SAGI, L. et al. Pemphigus and infectious agents. **Autoimmunity Reviews**, v. 8 (1), p. 33–35, 2008.

SALVIANO-SILVA, A.; PETZL-ERLER, M.L; BOLDT, A.B.W. CD59 polymorphisms are associated with gene expression and different sexual susceptibility to pemphigus foliaceus. **Autoimmunity**, v50(6), 377-385, 2017.

SALVIANO-SILVA, A. & LOBO-ALVES, S.C. et al. Besides Pathology: Long Non-Coding RNA in Cell and Tissue Homeostasis. **Non-coding RNA**, 4(1), 3, 2018.

SANTIN, I. et al. Celiac Disease–associated lncRNA Named HCG14 Regulates NOD1 Expression in Intestinal Cells. **Journal of Pediatric Gastroenterology and Nutrition**, 67(2), 225–231, 2018. doi:10.1097/mpg.0000000000001970

SARIG, O. et al. Population-specific association between a polymorphic variant in ST18, encoding a pro-apoptotic molecule, and pemphigus vulgaris. **J Invest Dermatol.**, 132:1798–805, 2012. 10.1038/jid.2012.46

SAXENA, A. et al. Preferential transmission of genetic risk variants of candidate loci at 6p21 from asymptomatic grandparents to mothers of children with neonatal lupus. **Arthritis and rheumatism**, 64(3), 931–939, 2012. doi:10.1002/art.33366

SCHMIDT, E.; KASPERKIEWICZ, M.; JOLY, P. Pemphigus. **Lancet** (London, England) Sep 7 394(10201):882–94, 2019. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31498102

SCHMIDT, E.; WASCHKE, J. Apoptosis in pemphigus. **Autoimmun Rev**., 8:533–7, 2009. doi: 10.1016/j.autrev.2009.01.011

SHI, X.et al. Long non-coding RNAs: A new frontier in the study of human diseases. **Cancer Letters**, v. 339(2), p.159 -166, 2013.

SHI, L. et al. The SLE transcriptome exhibits evidence of chronic endotoxin exposure and has widespread dysregulation of non-coding and coding RNAs. **PLoS One**, v.9, 2014.

SINHA, A.A. The Genetics of Pemphigus. **Dermatologic Clinics**, v. 29 (3), p. 381–391, 2011.

SMITH, M.A.; MATTICK, J. S. Structural and Functional Annotation of Long Noncoding RNAs. **Bioinformatics**, 65–85, 2016.

SONG J. et al. PBMC and exosome-derived Hotair is a critical regulator and potent marker for rheumatoid arthritis. **Clinical and Experimental Medicine**, v.15121–126, 2015.

SONKOLY, E. et al. Identification and Characterization of a Novel, Psoriasis Susceptibility-related Noncoding RNA gene, PRINS. **The Journal Of Biological Chemistry**, v. 280(25), p. 24159–24167, 2005.

SPADONI, M.B. et al. First Glimpse of Epigenetic Effects on Pemphigus Foliaceus. **J Invest Dermatol.**, Feb;140(2):488-491.e1,. doi: 10.1016/j.jid.2019.07.691, 2020. Epub 2019 Jul 31.

SPINDLER, V. et al. Mechanisms Causing Loss of Keratinocyte Cohesion in Pemphigus. Vol. 138, **Journal of Investigative Dermatology**. Elsevier B.V.; p. 32–7, 2018.

STUHLMULLER, B et al. Detection of oncofetal h19 RNA in rheumatoid arthritis synovial tissue. **Am J Pathol.**, v.163, p.901–911, 2003.

TAKEICHI, M. Cadherin cell adhesion receptors as a morphogenetic regulator. **Science**, v. 251 (5000), p. 1451-1455, 1991.

TANASILOVIC, S. et al. Pemphigus vulgaris and pemphigus foliaceus determined by CD86 and CTLA4 polymorphisms. **Clin Dermatol**., 35:236–41, 2017. 10.1016/j.clindermatol.2016.05.021

THAMI, P.K.; CHIMUSA, E.R. Population Structure and Implications on the Genetic Architecture of HIV-1 Phenotypes Within Southern Africa. **Frontiers in genetics**, 10, 905, 2019. doi:10.3389/fgene.2019.00905

THE RNACENTRAL CONSORTIUM. RNAcentral: a hub of information for non-coding RNA sequences. **Nucleic Acids Res.;** 47(D1):D221–D229, 2019. doi:10.1093/nar/gky1034

TOIBER, D.; LEPRIVIER, G.; & ROTBLAT, B. Long noncoding RNA: noncoding and not coded. **Cell Death Discovery**, 3, 16104, 2017. doi:10.1038/cddiscovery.2016.104

TRON, F. et al. Genetic factors in pemphigus. **J Autoimmun.**, Jun;24(4):319-28, 2005.

URAYAMA, K.Y. et al. Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. **Journal of the National Cancer Institute**, 104(3), 240–253, 2012. doi:10.1093/jnci/djr516

VODO, D.; SARIG, O.; & SPRECHER, E. The Genetics of Pemphigus Vulgaris. **Frontiers in medicine**, 5, 226, 2018. doi:10.3389/fmed.2018.00226

WANG, K.C.; CHANG, H.Y. Molecular Mechanisms of Long Noncoding RNAs. **Molecular Cell**, v. 43 (6), p.. 904–914, 2011.

WAPINSKI, O.; CHANG, H. Y. Long noncoding RNAs and human disease. **Trends in Cell Biology**, v. 21(6), p.354–361, 2011.

WEST, K.A.; LAGOS, D. Long Non-Coding RNA Function in CD4+ T Cells: What We Know and What Next? **Non-coding RNA**, 12 Jul 2019, 5(3). DOI: 10.3390/ncrna5030043

WILUSZ, J.E. et al. Long noncoding RNAs: functional surprises from the RNA world. **Genes Dev**., v. 23, p. 1494–1504, 2009.

XIE, W. et al. Genome-Wide Analyses Reveal Gene Influence on HIV Disease Progression and HIV-1C Acquisition in Southern Africa. **AIDS research and human retroviruses**, 33(6), 597–609, 2017. doi:10.1089/AID.2016.0017

YAN, L., WANG, J. M., & ZENG, K. Association between HLA-DRB1 polymorphisms and pemphigus vulgaris: a meta-analysis. **The British journal of dermatology**, 167(4), 768–777, 2012. doi:10.1111/j.1365-2133.2012.11040.x

YATAGAI, Y. et al. Variants near the HLA complex group 22 gene (HCG22) confer increased susceptibility to late-onset asthma in Japanese populations. **Journal of Allergy and Clinical Immunology**, 138(1), 281–283.e13, 2016. doi:10.1016/j.jaci.2015.11.023

ZHANG, H. et al. LncRNA HCG11 promotes proliferation and migration in gastric cancer via targeting miR-1276/CTNNB1 and activating Wnt signaling pathway. **Cancer cell international**, 19, 350, 2019. doi:10.1186/s12935-019-1046-0

ZHANG, L. et al. Long non-coding RNA HCG11 suppresses the growth of glioma by cooperating with the miR-4425/MTA3 axis. **J Gene Med**. Apr;21(4):e3074, 2019. doi: 10.1002/jgm.3074.

ZHANG, M. et al. A C6orf10/LOC101929163 locus is associated with age of onset in C9orf72 carriers. **Brain : a journal of neurology**, 141(10), 2895–2907, 2018. doi:10.1093/brain/awy238

ZHENG, Y. et al. Identification of potential hub-lncRNAs in ischemic stroke based on Subpathway-LNCE method. **Journal of Cellular Biochemistry**, 2019. doi:10.1002/jcb.28554

ZILIOTTO, N. et al. C6orf10 Low-Frequency and Rare Variants in Italian Multiple Sclerosis Patients. **Frontiers in genetics**, 10, 573, 2019. doi:10.3389/fgene.2019.00573

## 9. TRABALHO ANEXO (Artigo de revisão de literatura publicado)

*non-coding*
*RNA*

*Review*

# Besides Pathology: Long Non-Coding RNA in Cell and Tissue Homeostasis

**Amanda Salviano-Silva †, Sara Cristina Lobo-Alves †, Rodrigo Coutinho de Almeida, Danielle Malheiros and Maria Luiza Petzl-Erler ***

Laboratory of Human Molecular Genetics, Department of Genetics, Universidade Federal do Paraná, Curitiba 81531-980, Caixa Postal 19071, Brazil; amandasalviano92@gmail.com (A.S.-S.); saralobo5@yahoo.com.br (S.C.L.-A.); rodrigocout@gmail.com (R.C.d.A.); mfdani@yahoo.com.br (D.M.)

* Correspondence: perler@ufpr.br or mlperler@gmail.com; Tel.: +55-41-3361-1746 or +55-41-3361-1724

† These joint-first authors contributed equally to this review.

Received: 30 December 2017; Accepted: 25 January 2018; Published: 30 January 2018

**Abstract:** A significant proportion of mammalian genomes corresponds to genes that transcribe long non-coding RNAs (lncRNAs). Throughout the last decade, the number of studies concerning the roles played by lncRNAs in different biological processes has increased considerably. This intense interest in lncRNAs has produced a major shift in our understanding of gene and genome regulation and structure. It became apparent that lncRNAs regulate gene expression through several mechanisms. These RNAs function as transcriptional or post-transcriptional regulators through binding to histone-modifying complexes, to DNA, to transcription factors and other DNA binding proteins, to RNA polymerase II, to mRNA, or through the modulation of microRNA or enzyme function. Often, the lncRNA transcription itself rather than the lncRNA product appears to be regulatory. In this review, we highlight studies identifying lncRNAs in the homeostasis of various cell and tissue types or demonstrating their effects in the expression of protein-coding or other non-coding RNA genes.

**Keywords:** long non-coding RNA; homeostasis; physiological regulatory mechanisms; gene expression; gene regulation; transcriptome

*Non-coding RNA* **2018**, *4*, 3; doi:10.3390/ncrna4010003 www.mdpi.com/journal/ncrna

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).